

PRO GRADU -TUTKIELMA

**Tarja Sirén**

**Yksimielisyyden mittaamisen tilastolliset menetelmät**

TAMPEREEN YLIOPISTO  
Matematiikan ja tilastotieteen laitos  
Tilastotiede  
Syyskuu 2008

Tampereen yliopisto  
Matematiikan ja tilastotieteen laitos  
SIRÉN, TARJA: Tutkielman nimi  
Pro gradu -tutkielma, 54 s., 2 liites.  
Tilastotiede  
Syyskuu 2008

---

## Tiivistelmä

Arvioijien yksimielisyyden mittaaminen on monen tutkimusalan peruskysymyksiä. Esimerkiksi teollisen tuotantoprosessin kehittämisessä voidaan olla kiinnostuneita siitä, antaako uusi tehokkaampi menetelmä yhtä hyviä tuloksia kuin vanha menetelmä. Tässä tutkimuksessa rajoitutaan tarkastelemaan tilanteita, joissa kaksi arvioijaa luokittelee samat havainnot toisistaan riippumattomasti kahteen luokkaan. Tutkimuksessa keskitytään Cohenin kappaan, joka on yksi tunnetuimmista ja sovelletuimmista yksimielisyyden indeksimitoista (Cohen 1960). Kuitenkin voidaan näyttää, että joissain erikoistilanteissa se käyttäytyy epäjohdonmukaisesti, mistä syystä sille on esitetty monia vaihtoehtoja.

Yhtä Cohenin kappan muunnosta kutsutaan modifoiduksi kapaksi. Se parantaa Cohenin kappan käyttäytymistä täydellisen erimielisyyden tapauksessa. Tutkimuksessa osoitetaan, että modifioitu kappaa on itseisarvoltaan suurempi tai yhtä suuri kuin Cohenin kappaa. Modifioitu kappaa yliarvioi siis Cohenin kappan arvon. Voidaan kuitenkin näyttää esimerkiksi simulointikokeiden avulla, että näiden kappojen erotus jää käytännön sovelluksissa useimmiten pieneksi. Edelleen on kuitenkin tilanteita, joissa Cohenin kappaa ja modifioitu kappaa käyttäytyvät epäjohdonmukaisesti. Tästä syystä tutkimuksessa esitellään vielä kaksi muunnosta, jotka perustuvat Cohenin kappan tai modifioidun kappan standardointiin, kun jokin tekijä on vakioitu. Yksityiskohtaisen tarkastelun tuloksena voidaan kuitenkin osoittaa yllättävä tulos, että vakioinnin seurauksena standardoidut kappat eivät enää mittaa yksimielisyyttä.

Koska muunnokset eivät tarjoa yksikäsitteisesti parempaa vastinetta Cohenin kappalle, otetaan tutkimuksessa kantaa siihen, miten Cohenin kappaa tulisi soveltaa käytännössä. Ensinnäkin otoskoko tulisi määrittää siten, että testauksessa ja väliestimoinnissa saavutetaan haluttu tarkkuus. Riittävä otoskoko voidaan määrittää ennen tutkimuksen suorittamista Cohenin kappan suurten otosten hajonnan maksimointiin perustuvalla menetelmällä. Toiseksi Cohenin kappan tulkinnassa tulee huomioida muun muassa arvioijien ammattitaito, kontingenssitaulukon rakenne ja hajonta.

**Asiasanat** Cohenin kappaa, modifioitu kappaa, havaittu yksimielisyys, sattumayksimielisyys, esiintyvyys, harha

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>5</b>
1.1	Taustaa ja sovelluskohteita . . . . .	5
1.2	Tutkielman tavoitteet . . . . .	6
1.3	Tutkielman rakenne . . . . .	6
<b>2</b>	<b>Yksimielisyyden mittaaminen</b>	<b>8</b>
2.1	Perusmääritelmät . . . . .	8
2.2	Cohenin kapin tulkinta . . . . .	11
2.3	Ongelmatilanteita . . . . .	12
<b>3</b>	<b>Cohenin kapin muunnoksia</b>	<b>18</b>
3.1	Modifioitu kappi . . . . .	18
3.2	Standardoidut kappit . . . . .	22
3.2.1	Havaittu yksimielisuus vakioitu . . . . .	22
3.2.2	Sattumayksimielisuus vakioitu . . . . .	25
3.3	Pohdintaa ja vertailua . . . . .	28
<b>4</b>	<b>Tilastollinen päättely yksimielisyyden mittaamisessa</b>	<b>32</b>
4.1	Binomi- ja multinomijakauma . . . . .	32
4.2	Taylorin sarjat . . . . .	34
4.3	Harhattomuus ja suurten otosten varianssi . . . . .	35
4.4	Testaus . . . . .	38
<b>5</b>	<b>Otoskoon määrittäminen</b>	<b>41</b>
5.1	Luottamusväli perusteinen . . . . .	41
5.2	Testaus perusteinen . . . . .	42
<b>6</b>	<b>Simulointi</b>	<b>44</b>
6.1	Bootstrap-menetelmä . . . . .	44
6.1.1	Teoria . . . . .	44
6.1.2	Simulointikokeiden perusidea . . . . .	45
6.2	Kappojen vertailu . . . . .	46
6.3	Hajontojen vertailu . . . . .	48
<b>7</b>	<b>Yhteenveto</b>	<b>51</b>
	<b>Lähdeluettelo</b>	<b>53</b>

Liite A: Maksimaalinen $Q$ , kun Cohenin kappa tuntematon	55
Liite B: Parametrin $Q$ maksimi, kun Cohenin kappan arvo arvioitu	56

# 1 Johdanto

## 1.1 Taustaa ja sovelluskohteita

Monilla tieteenaloilla on sovelluksia, joissa halutaan arvioida yksimielisyyttä. Ollaan esimerkiksi kiinnostuneita useamman lääkärin tai psykologin yksimielisyydestä, kun he diagnosoivat samat potilaat. Tai halutaan tutkia, tuottaako uusi edullisempi menetelmä tai laite samanlaisia tuloksia kuin hyvänä mittaapuuna pidetty menetelmä. Yleisesti ottaen yksimielisyyden mittaamista sovelletaan kaikissa tilanteissa, joissa eri arvioijat arvioivat samat havainnot. Käytännön sovelluksia on siis todella paljon monilla tieteenaloilla. Onkin mielenkiintoista tutkia, minkälaisia tilastollisia menetelmiä yksimielisyyden mittaamiseksi käytännössä hyödynnetään ja kuinka luotettavasti ne toimivat erilaisissa tilanteissa.

Ensimmäisen yksimielisyyttä mittaavan tunnusluvun esittivät Goodman & Kruskal (1954). Menetelmä perustuu havaittuun yksimielisyyteen, mutta se ei huomioi sitä, että tietty osa havaitusta yksimielisyydestä selittyy sattumalla. Tunnetuin käytössä oleva sattumayksimielisyyden huomioiva indeksimitta on Jacob Cohenin vuonna 1960 esittämä Cohenin kappa (Cohen 1960). Se soveltuu tilanteisiin, joissa kaksi arvioijaa luokittelee toisistaan riippumattomasti samat havainnot kahteen tai useampaan nominaaliasteikolliseen luokkaan. Vastaavanlainen yksimielisyyden mittari on muun muassa Blochin ja Kraemerin (1989) esittämä menetelmä, joka rajoittuu tilanteisiin, joissa oletetaan, että kumpikin arvioija asettaa havainnon tiettyyn luokkaan samalla todennäköisyydellä. (Banerjee, Capozzoli, McSweeney & Sinha, 1999.) Myös log-lineaarisia malleja on sovellettu yksimielisyyttä mitattaessa (Agresti 2007). Yksimielisyyden mittaaminen Cohenin kappan avulla on yleistetty myös tapauksiin, joissa luokittelu on järjestysasteikollinen (Cohen 1968). Tätä menetelmää kutsutaan painotetuksi kapaksi. Muun muassa Fleiss (1971) ja Light (1971) ovat esitelleet yleistyksen myös useamman arvioijan tilanteeseen.

Paljon muitakin yleistyksiä ja niiden tilastollisia ominaisuuksia on tutkittu vuosien varrella. Tässä tutkimuksessa keskitytään tarkastelemaan Cohenin kappa, sillä se on tunnetuin ja yleisesti käytetty yksimielisyyden tunnusluku. Tarkasteluissa rajoitutaan tilanteisiin, joissa kaksi arvioijaa luokittelee  $n$  riippumattomaa havaintoa täysin itsenäisesti yhteen kahdesta nominaaliasteikollisesta luokasta. Cohenin kappan ominaisuuksia tutkiessa voidaan huomata, että sen tulkinta vääristyy jossain erikoistilanteissa. Sille on esitetty useita erilaisia muunnoksia, jotta tämä epäjohdonmukainen käyttäytyminen saataisiin korjat-

tua. Kuitenkin useimmissa muunnoksissa ilmenee uusia ongelmia tai rajoitteita, mistä syystä Cohenin kappaa on säilyttänyt asemansa yksimielisyyden mittana käytännössä.

Tässä tutkimuksessa tuodaan esille muutamia melko tuoreita ja siten vähän tutkittuja muunnoksia Cohenin kapalle ja tutkitaan niiden toimivuutta kriittisesti. Lisäksi tutkitaan Cohenin kappaa ja varteenotettavien muunnosten tilastollisia ominaisuuksia. Muun muassa suurten otosten varianssin tunteminen mahdollistaa otoskoon määrittämisen ennen tutkimuksen suorittamista, kun kiinnitetään testauksen tarkkuus tai luottamusvälin maksimaalinen leveys. Käytännössä tämä on osoittautunut todella hyödylliseksi, jotta tutkimuksen resurssit voidaan kohdistaa tehokkaasti.

## 1.2 Tutkielman tavoitteet

Tutkielman tavoitteena on esitellä Cohenin kappaa ja siihen perustuvia muunnoksia yksityiskohtaisesti siten, että lukija saa laajan käsityksen niiden keskeisistä ominaisuuksista ja, että hänen on helppo sisäistää, mihin muunnokset perustuvat ja mikä niiden hyöty ja tulkinta on käytännössä. Tämän tavoitteen saavuttamiseksi apuna käytetään pohdinnan ja matemaattisen tarkastelun lisäksi simulointikokeita, joiden avulla pyritään havainnollistamaan kappojen ominaisuuksia vielä tarkemmin. Tutkimuksessa rajoitetaan tutkimaan kahden arvioijan ja kahden päätösvaihtoehdon tapausta, sillä yksinkertaisessa tapauksessa tunnuslukujen epä johdonmukaisen käyttäytymisen syyt ja ominaisuudet on helpompi havaita ja perustella. Tältä pohjalta mielenkiintoisiksi nousevat tulokset voidaan jatkossa laajentaa yleisempiin tilanteisiin.

Kriittisen tarkastelun tuloksena tutkimuksessa käy ilmi, etteivät esitetyt muunnokset tarjoa yksikäsitteistä ja kaikissa tilanteissa johdonmukaisesti toimivaa yksimielisyyden tunnuslukua. Siksi erityistä tarkkuutta tulee kiinnittää tulkintaan. Tutkimuksen tarkoituksena onkin ottaa kantaa erityisesti Cohenin kappaa tulkintaan ja pohtia, mitkä tekijät tulee huomioida ja miten tulkintaa tehtäessä.

Kokonaisuudessaan tutkimuksen tavoitteena ei ole listata kaikkia yksimielisyyttä mittaavia tunnuslukuja tai menetelmiä, vaan tarkoituksena on tutkia kriittisesti ja yksityiskohtaisesti Cohenin kappaa ja sille perustuvia muutamia melko tuoreita ja vähän tutkittuja muunnoksia. Tavoitteena on siis tunnistaa yleisesti käytössä olevan tunnusluvun epä johdonmukaisuudet ja pyrkiä muuttamaan sitä ongelmatilanteiden pohjalta.

## 1.3 Tutkielman rakenne

Johdannon jälkeen luvussa 2 esitellään Cohenin kappaa ja yksimielisyyden mittaamiseen liittyviä keskeisiä käsitteitä. Lisäksi tutkitaan yksityiskohtaisesti Cohenin kappaa ominaisuuksia ja tulkintaa. Tämän jälkeen keskitytään muunnoksiin, joiden tavoitteena on parantaa Cohenin kappaa tulkintaa tilanteissa, joissa

se käyttäytyy epäjohdonmukaisesti (luku 3, s. 18). Näitä muunnoksia tarkastellaan kriittisesti ja pohditaan, helpottavatko ne yksimielisyyden tason tulkintaa. Luvussa 4 esitellään Cohenin kapan ja modifioidun kapan tilastollisia ominaisuuksia kuten odotusarvo, suurten otosten varianssi ja testaus (s. 32). Tämän jälkeen luvussa 5 keskitytään otoskoon määrittämiseen. Näin varmistetaan, että testauksessa ja väliestimoinnissa saavutetaan haluttu tarkkuus (s. 41). Viimeisenä ennen yhteenvetoa arvioidaan kappojen käyttäytymistä ja ominaisuuksia simulointikokeiden avulla (luku 6, s. 44). Liitteissä A ja B esitellään taulukot Cohenin kapan maksimaalisista hajonnoista erilaisilla taustaoletuksilla. Näitä taulukoita hyödynnetään otoskokoa määrittäessä.

Läpi tutkielman teorian rinnalla tuloksia havainnollistetaan käyttäytymistieteisiin liittyvällä aineistolla. Aineisto on peräisin kyselytutkimuksesta, joka suoritettiin neljällä Helsingin asuinalueella (Eestinkallio-Puolarniitty, Kumpula-Toukola, Latokoski ja Näkinpuisto). Tarkempia tietoja aineistosta ja tutkimustuloksista löytyy Korpelan ja Ylénin (2007) artikkelista.

Tutkielman laskennalliset tarkastelut on suoritettu R 2.7.1 –ohjelmistolla.

## 2 Yksimielisyyden mittaaminen

On paljon käytännön sovelluksia, joissa halutaan mitata arvioijien yksimielisyyttä tai esimerkiksi kahden menetelmän tai lääkevaikutuksen samankaltaisuutta. Eräs karkea tapa arvioida yksimielisyyden tasoa on laskea prosenttiosuus niistä tilanteista, joissa arvioijat ovat samaa mieltä. Tämä tapa ei kuitenkaan huomioi sitä todennäköisyyttä, jolloin arvioijat antavat saman arvion täysin sattumalta. Tästä syystä yksimielisyyden mittaamista varten on kehitelty menetelmiä, joissa tavoitteena on minimoida sattumalta syntyneen yksimielisyyden aiheuttama harha. Yksi tunnetuimmista harhan huomioivista yksimielisyyden mittareista on Cohenin kappa (Cohen 1960). Se soveltuu tilanteisiin, joissa päätöksiä on tekemässä kaksi arvioijaa ja päätösvaihtoehdot ovat luokitteluasteikollisia.

### 2.1 Perusmääritelmät

Tutkitaan tilannetta, jossa arvioijat A ja B luokittelevat täysin itsenäisesti  $n$  tilastollisesti riippumatonta yksilöä yhteen kahdesta mahdollisesta luokasta. Luokat ovat toisensa poissulkevia, tyhjentäviä ja nominaaliasteikollisia. Merkitään näitä luokkia 1:llä ja 2:lla. Tilanne on esitetty taulukossa 2.1. Tässä taulukossa  $\pi_{ij}$  on arvioiden suhteellinen osuus, missä  $i$  viittaa arvioijaan A ja  $j$  arvioijaan B ( $i, j = 1, 2$ ). Suhteellinen osuus kuvaa arvioijien yksimielisyyttä, kun indeksit ovat yhtä suuret ( $i = j$ ). Erimielisyyttä sen sijaan kuvaavat suhteelliset osuudet, joissa  $i \neq j$ . Rivisummat

$$\pi_{1\cdot} = \sum_{j=1}^2 \pi_{1j} \quad \text{ja} \quad \pi_{2\cdot} = \sum_{j=1}^2 \pi_{2j}$$

ovat sellaisia suhteellisia osuuksia, joilla arvioija A luokittelee yksilön luokkaan 1 tai 2. Merkinnöillä  $\pi_{1\cdot}$  ja  $\pi_{2\cdot}$  viitataan vastaaviin arvioijan B antamiin arvioihin.

Cohenin kappa määritellään nyt seuraavasti (Cohen 1960)

$$(2.1) \quad \kappa_C = \frac{\theta_0 - \theta_E}{1 - \theta_E},$$

missä  $\theta_0 = \pi_{11} + \pi_{22}$  kertoo suhteellisen osuuden, jolloin arvioijat ovat yksimielisiä. Sen sijaan  $\theta_E = \pi_{1\cdot}\pi_{\cdot 1} + \pi_{2\cdot}\pi_{\cdot 2}$  kuvaa sattumayksimielisyyden suuruutta, missä oletetaan, että arvioijien päätökset ovat riippumattomia.



**Taulukko 2.1.** Kontingenssitaulukko kahden arvioijan tilanteessa.

		Arvioija B		
		1	2	
Arvioija A	1	$\pi_{11}$	$\pi_{12}$	$\pi_{1\cdot}$
	2	$\pi_{21}$	$\pi_{22}$	$\pi_{2\cdot}$
		$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	

Sanallisesti ilmaistuna kapan arvo on havaitun yksimielisyyden ja mahdollisen yksimielisyyden välinen suhde, kun sattumayksimielisyys huomioidaan. Cohenin kapan arvo perustuu havaitun yksimielisyyden ja sattumayksimielisyyden erotukseen. Karkeasti ottaen, mitä suurempi ero kyseisten suhteellisten osuuksien välillä on, sitä suurempi Cohenin kapan arvo on itseisarvoltaan. Cohenin kapan vaihteluväli on  $[-1, 1]$ .

Yllä esitetyssä tarkastelussa oletetaan, että populaation suhteelliset osuudet ovat tunnettuja. Näin ei kuitenkaan useimmiten ole, vaan suhteelliset osuudet joudutaan estimoimaan otoksen perusteella. Käytännössä Cohenin kappa estimoidaan otossuurella

$$(2.2) \quad \hat{\kappa}_C = \frac{\hat{\theta}_0 - \hat{\theta}_E}{1 - \hat{\theta}_E},$$

missä tuntemattomat populaation parametrit  $\pi_{ij}$ ,  $\pi_{i\cdot}$  ja  $\pi_{\cdot j}$  korvataan vastaavilla otoksesta lasketuilla suureilla  $\hat{\pi}_{ij} = \frac{n_{ij}}{n}$ ,  $\hat{\pi}_{i\cdot} = \frac{n_{i\cdot}}{n}$ ,  $\hat{\pi}_{\cdot j} = \frac{n_{\cdot j}}{n}$ ,  $\hat{\theta}_0 = \frac{n_{11} + n_{22}}{n}$  ja  $\hat{\theta}_E = \frac{n_{1\cdot}n_{\cdot 1} + n_{2\cdot}n_{\cdot 2}}{n^2}$  (Cohen 1960).

Määritellään edelleen esiintyvyys ja harha, joita tullaan jatkossa hyödyntämään Cohenin kapan tulkinnan yhteydessä. Esiintyvyys määritellään diagonaalialkioiden erotuksena

$$(2.3) \quad \hat{\theta}_P = |\hat{\pi}_{11} - \hat{\pi}_{22}| = \frac{|n_{11} - n_{22}|}{n}.$$

Sen sijaan harha on ei-diagonaalialkioiden erotus

$$(2.4) \quad \hat{\theta}_B = |\hat{\pi}_{12} - \hat{\pi}_{21}| = \frac{|n_{12} - n_{21}|}{n}.$$

(Sim & Wright 2005.) Tulee kuitenkin huomata, ettei harhalla viitata tilastolliseen harhaan, jolla tarkoitetaan piste-estimaattorin odotusarvon ja estimoitavan parametrin erotusta (Casella & Berger 2002). Havainnollistetaan seuraavassa esimerkissä Cohenin kapan, esiintyvyyden ja harhan estimointia käytännön aineiston avulla.

**Esimerkki 2.1.** Esimerkkiaineisto on peräisin kyselytutkimuksesta, jossa tutkittiin vastaajan terveydentilan, mielialan valinnan sekä kokemusten yhteyttä (Korpela et al. 2007). Tutkimuksessa vastaajilta kysyttiin muun muassa mielialaa ja toiminnallisuutta siellä. Mielipaikka tuli kuvailla, merkitä karttaan ja kertoa miksi se oli mieluinen.

Mielipaikat luokiteltiin kahdeksaan luokkaan kahden riippumattoman arvioijan toimesta. Arvioijille annettiin luokitteluperusteet ja he tekivät päätöksensä toisistaan riippumattomasti ensisijaisesti paikan kuvailun perusteella. Ainoastaan tarvittaessa he käyttivät apunaan karttaa tai vastaajan perustelua mielipaikan valinnalle.

Ollaan kiinnostuneita siitä, ovatko arvioijat luokittelleet vastaajien mielipaikat yksimielisesti luonnon (muun muassa puistot, metsät, rannat) ja muiden paikkojen välillä. Mielipaikkamuuttuja on siis luokiteltu uudestaan siten, että luonto muodostaa oman ja muut paikat toisen luokan. Yksimielisyyden voimakkuutta voidaan mitata Cohenin kapin avulla, sillä havainnot kuten myös arvioijien tekemät luokittelut ovat toisistaan riippumattomia.

**Taulukko 2.2.** Frekvenssitaulukko kahden arvioijan mielipaikkaluokittelusta.

	Luonto	Muu	
Luonto	37	4	$n_{1.} = 41$
Muu	5	48	$n_{2.} = 53$
	$n_{.1} = 42$	$n_{.2} = 52$	$n = 94$

Taulukon 2.2 perusteella voidaan estimoida esiintyvyys  $\hat{\theta}_P$ , harha  $\hat{\theta}_B$  ja Cohenin kapin arvo  $\hat{\kappa}_C$  (lausekkeet 2.3, 2.4 ja 2.2). Tästä saadaan

$$\begin{aligned}\hat{\theta}_0 &= \hat{\pi}_{11} + \hat{\pi}_{22} = \frac{37}{94} + \frac{48}{94} \approx 0.904, \\ \hat{\theta}_E &= \hat{\pi}_{1.}\hat{\pi}_{.1} + \hat{\pi}_{2.}\hat{\pi}_{.2} = \left(\frac{41}{94} \times \frac{42}{94}\right) + \left(\frac{53}{94} \times \frac{52}{94}\right) \approx 0.507, \\ \hat{\theta}_P &= |\hat{\pi}_{11} - \hat{\pi}_{22}| = \left|\frac{37}{94} - \frac{48}{94}\right| \approx 0.117, \\ \hat{\theta}_B &= |\hat{\pi}_{12} - \hat{\pi}_{21}| = \left|\frac{4}{94} - \frac{5}{94}\right| \approx 0.011,\end{aligned}$$

Havaittu yksimielisyys on todella suuri  $\hat{\theta}_0 = 0.904$ , kun taas sattumayksimielisyys  $\hat{\theta}_E = 0.507$ . Esiintyvyys ja harha ovat kummatkin melko pieniä, mikä kertoo siitä, että niin eriävät kuin myös yksimieliset vastaukset jakautuvat melko tasaisesti mielipaikkaluokkien suhteen. Estimoidaan edelleen arvioijien marginaalisummat.

$$\begin{aligned}\hat{\pi}_{1.} &= \frac{37}{94} + \frac{4}{94} \approx 0.436, \\ \hat{\pi}_{.1} &= \frac{37}{94} + \frac{5}{94} \approx 0.447, \\ \hat{\pi}_{2.} &= \frac{5}{94} + \frac{48}{94} \approx 0.564, \\ \hat{\pi}_{.2} &= \frac{4}{94} + \frac{48}{94} \approx 0.553 \text{ ja}\end{aligned}$$

$$\hat{\kappa}_C = \frac{\hat{\theta}_0 - \hat{\theta}_E}{1 - \hat{\theta}_E} \approx \frac{0.904 - 0.507}{1 - 0.507} = 0.806.$$

Havaittu Cohenin kappan arvo on melko suuri, joten voidaan päätellä, että arvioijien välinen yksimielisyys on hyvä. Jos jatkotutkimuksessa halutaan hyödyntää mielipaikkamuuttujaa erilaisissa tilastollisissa riippuvuustarkasteluissa, tulee tutkijoiden päättää, miten eriävien mielipiteiden kohdalla toimitaan. Yksi mahdollisuus on käyttää vain toisen arvioijan tuloksia. Toisaalta, jos eriäviä vastauksia ei ole paljon, voitaisiin myös tutkia nämä tilanteet erikseen ja valita luokka tapauskohtaisesti.

Esimerkin tapauksessa Cohenin kappan vaikutti toimivan johdonmukaisesti, sillä sen avulla tehty päätelmä yksimielisyyden tasosta tukee frekvenssitaulukon rakennetta. Seuraavassa alaluvussa tarkastellaan vielä tarkemmin Cohenin kappan tulkintaa.

## 2.2 Cohenin kappan tulkinta

Cohenin kappan voi saada arvoja välillä  $[-1, 1]$ . Täydellistä yksimielisyyttä mittaa arvo 1, kun taas Cohenin kappan arvo 0 vastaa tilannetta, jossa arvioijat olisivat arvanneet kaikki tapaukset. Negatiivinen Cohenin kappan arvo kertoo huonommasta kuin sattumanvaraisesta yksimielisyydestä. Kuitenkin useimmissa tarkasteluissa on harvinaista saada kappan arvoksi negatiivinen luku. Taulukossa 2.3 on Landisin ja Kochin (1977) ehdottama asteikko Cohenin kappan voimakkuuden arvioimiseksi. Asteikon perusteella voidaan sanoa, että yksimielisyys arvioijien välillä on todella hyvä, jos  $\kappa_C > 0.8$ .

**Taulukko 2.3.** Asteikko Cohenin kappan voimakkuudelle (Landis et al. 1977).

Cohenin kappan	Yksimielisyyden voimakkuus
<0.00	Huono
0.00–0.20	Heikko
0.21–0.40	Kohtalainen
0.41–0.60	Keskinkertainen
0.61–0.80	Huomattava
0.81–1.00	Lähes täydellinen

Kuitenkin asteikko on vain karkea arvio Cohenin kappan suuruuden mittaamiseksi, sillä voidaan ajatella, että muun muassa tutkimuksen kohteena olevan aiheen vakavuus vaikuttaa rajan asettamiseen merkittävästi. Tämän lisäksi Sim et al. (2005) esittivät, että kappan suuruuteen ja siten myös sen tulkintaan vaikuttavat esiintyvyys, harha ja päätösten riippuvuus.

Päätökset ovat riippuvia tilanteissa, joissa arvioijien tekemät päätökset tai arvioitavat yksilöt eivät ole riippumattomia. Vastaavissa tapauksissa Cohenin

kapan arvo vääristyy. Täten Cohenin kappa ei sovellu esimerkiksi tilanteeseen, jossa toisen arvioijan tulisi vahvistaa toisen arvioijan tekemät lausunnot. Tarkastellaan seuraavassa alaluvussa vielä laajemmin muita tilanteita ja tekijöitä, jotka aiheuttavat Cohenin kapan arvon vääristymisen tai tulkinnan hankaloitumisen.

### 2.3 Ongelmatilanteita

Tässä alaluvussa tarkastellaan tarkemmin tilanteita, joissa Cohenin kappa ei käyttyädy johdonmukaisesti. Aluksi tutkitaan tilannetta, jossa ainakin toinen arvioija asettaa kaikki havainnot vain yhteen luokkaan. Näissä tilanteissa Cohenin kappa saa aina arvon 0. Taulukossa 2.4 havainnollistetaan kontingenssitaulukon rakennetta yhdessä vastaavanlaisessa tilanteessa.

**Lause 2.1.** *Jos jokin  $\pi_{.j} = 0$  tai  $\pi_{i.} = 0$ , niin Cohenin kappa  $\kappa_C = 0$  ( $i, j = 1, 2$ ).*

*Todistus.* Olkoon todennäköisyys sille, että arvioija  $X (= A, B)$  asettaa kaikki arviot samaan luokkaan, 1. Merkitään sellaista todennäköisyyttä, että arvioijat ovat samaa mieltä, parametrilla  $\alpha$ . Nyt saadaan, että toisen luokan marginaalisummien on oltava 1 ja  $\alpha$  ja toisen 0 ja  $1 - \alpha$ . Siten  $\theta_E = 1 \times \alpha + 0 \times (1 - \alpha) = \alpha = \theta_0$ , joten  $\kappa_C = 0$ . Tämä toteutuu aina, jos jokin marginaalisumma on 0  $\Leftrightarrow$  jokin marginaalisumma on 1.  $\square$

**Taulukko 2.4.** Kontingenssitaulukko, kun arvioija A asettaa kaikki havainnot luokkaan 1 ( $0 \leq \alpha \leq 1$ ).

		Arvioija B		
		1	2	
Arvioija A	1	$\alpha$	$1 - \alpha$	1
	2	0	0	0
		$\alpha$	$1 - \alpha$	1.0

Lauseen 2.1 perusteella voi syntyä tilanteita, joissa Cohenin kappa  $\kappa_C = 0$ , vaikka arvioijat olisivat yksimielisiä kaikkien havaintojen suhteen ( $\alpha = 1$ ) tai lähes kaikkien havaintojen suhteen (esim.  $\alpha = 0.95$ ). Tulos voidaan perustella sillä, että jos toinen arvioija asettaa kaikki havainnot yhteen luokkaan todennäköisyydellä 1, selittyy havaittu yksimielisyys täysin sattumalla. Näin ollen Cohenin kappa ei anna mielekästä tulkintaa, jos ainakin toisen arvioijan mielestä arvioitavat havainnot voidaan luokitella vain yhteen luokkaan.

Tarkastellaan lisäksi tarkemmin tilannetta, jossa Cohenin kappa  $\kappa_C = -1$ . Toivottua olisi, että Cohenin kappa saa arvon  $-1$  aina, kun havaittu yksimielisyys  $\theta_0 = 0$ . Tällöin Cohenin kappa olisi  $-1$  kaikissa taulukon 2.5 tapauksissa.

**Taulukko 2.5.** Kontingenssitaulukko täydellisen erimielisyyden tapauksessa. ( $0 < \beta < 1$ )

		Arvioija B		
		1	2	
Arvioija A	1	0	$\beta$	$\beta$
	2	$1 - \beta$	0	$1 - \beta$
		$1 - \beta$	$\beta$	

Voidaan kuitenkin osoittaa, että näin tapahtuu, jos ja vain jos parametri  $\beta = 0.50$ . Näin ollen esimerkiksi tilanteessa, jossa parametri  $\beta = 0.75$  on Cohenin kappaa  $\kappa_C = -0.60 \neq -1$ . Toisin sanoen Cohenin kappaa ei pysty kuvaamaan täydellistä erimielisyyttä muulloin kuin yksittäisessä erikoistapauksessa  $\beta = 0.50$ .

**Lause 2.2.**  $\kappa_C = -1$ , jos ja vain jos  $\pi_{12} = \pi_{21} = 0.50$ .

*Todistus.* Oletetaan, että  $\kappa_C = -1$ . Täydellisen erimielisyyden tapauksessa  $\theta_0 = 0$ , joten kontingenssitaulukko on muotoa 2.5. Näin ollen identiteetit

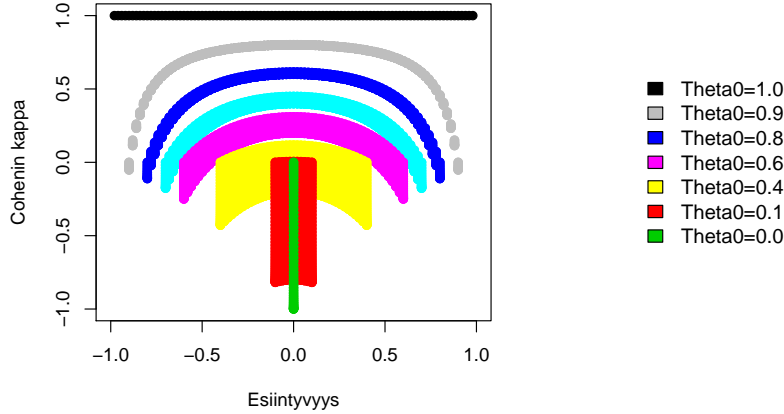
$$\begin{aligned} \frac{\theta_0 - \theta_E}{1 - \theta_E} &= -1 \\ \theta_0 - \theta_E &= \theta_E - 1 \\ 1 &= 2\theta_E \\ 1 &= 2(\pi_{1.}\pi_{.1} + \pi_{2.}\pi_{.2}) \\ 1 &= 2(\pi_{11} + \pi_{12})(\pi_{11} + \pi_{21}) + (\pi_{12} + \pi_{22})(\pi_{21} + \pi_{22}) \\ 1 &= 2(\pi_{12}\pi_{21} + \pi_{12}\pi_{21}) \\ 1 &= 4 \times \pi_{12}\pi_{21} \\ \pi_{12}\pi_{21} &= 1/4, \end{aligned}$$

ovat yksikäsitteisiä. Viimeisin identiteetti toteutuu jos ja vain jos  $\pi_{12} = \pi_{21} = 0.5$ .

Oletetaan seuraavaksi, että  $\pi_{12} = \pi_{21} = 0.50$ . Näin ollen  $\pi_{11} = \pi_{22} = 0$ , sillä  $\sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} = 1$ . Täten suoraviivaisesti laskemalla saadaan, että Cohenin kappaa  $\kappa_C = -1$ .  $\square$

Sim et al. (2005) esittivät, että Cohenin kappaa tulkitaan vaikuttavat muun muassa esiintyvyys ja harha, joten tutkitaan näiden tunnuslukujen vaikutusta Cohenin kappaan vielä tarkemmin. Esiintyvyyden vaikutus Cohenin kappaan realisoituu tilanteissa, joissa diagonaalialkiot  $\pi_{11}$  ja  $\pi_{22}$  eroavat toisistaan huomattavasti. Tällaiseen tilanteeseen voidaan joutua muun muassa silloin, kun toinen luokka kuvaa todella harvinaista ominaisuutta, sairautta tai tapahtumaa (lievempi tapaus lauseen 2.1 tilanteesta). Toisin sanoen päätösvaihtoehtojen jakauma on vino. Kuviossa 2.1 havainnollistetaan Cohenin kappaa riippuvuutta esiintyvyydestä havaitun yksimielisyyden  $\theta_0$  eri arvoilla. Kuvaaajassa

esiintyvyys on laskettu ilman itseisarvoa ja eri väreillä viitataan havaitun yksimielisyyden eri tasoihin.



**Kuvio 2.1.** Cohenin kappan riippuvuus esiintyvyydestä eri parametrin  $\theta_0$  arvoilla. (Esiintyvyys on laskettu ilman itseisarvoa.)

Kuvaajasta havaitaan, että Cohenin kappan minimi vaihtelee voimakkaasti eri esiintyvyyden tasoilla maksimin pysyessä vakiona. Minimien vaihtelu esiintyvyyden funktiona selittyy sillä, että esiintyvyyden ollessa suuri myös havaitun yksimielisyyden tulee olla suuri, joka taas vaikuttaa kappan arvoon. Maksimi sen sijaan on kaikilla esiintyvyyden tasoilla 1, sillä millä tahansa esiintyvyyden tasolla diagonaalialkiot voidaan muodostaa siten, että niiden summa on 1. Toisaalta on helppo näyttää, että Cohenin kappa saa arvon 1, jos ja vain jos diagonaalialkioiden summa on 1 eli havaittu yksimielisyyden  $\theta_0 = 1$ . Kuvaajan avulla vahvistuu myös tulos, jonka mukaan Cohenin kappa saa arvon  $-1$  vain yksittäisessä erikoistapauksessa, sillä havaitun yksimielisyyden ollessa nollassa, Cohenin kappan arvot vaihtelevat välillä  $[-1, 0]$  (lause 2.2).

Mielenkiintoinen havainto on, että tietyllä havaitun yksimielisyyden tasolla  $\theta_0$  kappan arvot pysyvät kutakuinkin samalla tasolla, mutta tietyn esiintyvyyden tason jälkeen kappojen arvot alkavat laskea jyrkästi. Suurilla havaitun yksimielisyyden  $\theta_0$  arvoilla lasku alkaa myöhemmin mutta on sitäkin jyrkempää, kun taas pienillä  $\theta_0$  arvoilla lasku alkaa aikaisemmin, mutta on loivempaa. Kaltevuuden muuttuminen havaitun yksimielisyyden  $\theta_0$  funktiona selittyy sillä, että suurilla havaitun yksimielisyyden tasoilla päätösvaihtoehtojen jakaumasta voi syntyä vinompi kuin heikommilla havaitun yksimielisyyden tasoilla. Toisaalta taas vinojen jakaumien tapauksessa sattumayksimielisyydestä  $\theta_E$  muodostuu suuri luku, mistä syystä havaittu yksimielisyyden selittyy lähes kokonaan sattumalla. Tästä syystä Cohenin kappan arvosta muodostuu näissä tilanteissa todella pieni suhteessa havaitun yksimielisyyden tasoon. Kuvaajan perusteella voidaan karkeasti arvioida, että esiintyvyyden ollessa yli 0.5 alkaa luokkien jakauman vinous vaikuttaa Cohenin kappan suuruuteen. Havainnollistetaan tilannetta vie-

lä esimerkin avulla.

**Esimerkki 2.2.** Tarkastellaan kahta kuvitteellista tilannetta, joissa kaksi lääkäriä tutkii 100 toisistaan riippumatonta potilasta ja antaa diagnoosin, onko potilas sairas vai ei. Lääkärit tutkivat samat sata potilasta, mutta tekevät päätökset toisistaan riippumattomasti.

Ensimmäisessä tilanteessa tutkittava sairaus on melko tavallinen ja otoksesakin on sairaita ja terveitä todennäköisesti lähestulkoon samassa suhteessa. Kuitenkaan ei tiedetä ketkä ovat terveitä ja ketkä sairaita. Kontingenssitaulukon 2.6 perusteella saadaan, että havaittu yksimielisyys lääkäreiden välillä on jopa 0.94 ja sattumayksimielisyys  $\hat{\theta}_E \approx 0.500$ . Täten estimoitu Cohenin kappa  $\hat{\kappa}_C \approx 0.880$ .

**Taulukko 2.6.** Kontingenssitaulukko lääkäreiden lausunnoista, kun sairaus on tavallinen.

	Terve	Sairas	
Terve	0.48	0.03	$\hat{\pi}_{1.} = 0.51$
Sairas	0.03	0.46	$\hat{\pi}_{2.} = 0.49$
	$\hat{\pi}_{.1} = 0.51$	$\hat{\pi}_{.2} = 0.49$	

Toisessa tapauksessa sairaus on harvinainen, joten myös otoksessa on todennäköisesti paljon terveitä. Vaikka havaittu yksimielisyys pysyykin samana ensimmäiseen tilanteeseen verrattuna, sen rakenne muuttuu siten, että  $\hat{\pi}_{11} = 0.93$  ja  $\hat{\pi}_{22} = 0.01$ . Myös ei-diagonaalialkiot ovat edelleen samat ( $\hat{\pi}_{12} = \hat{\pi}_{21} = 0.03$ ) (taulukko 2.7). Nyt havaittu sattumayksimielisyys  $\hat{\theta}_E \approx 0.923$ , joten estimoitu Cohenin kappa  $\hat{\kappa}_C \approx 0.219$ .

**Taulukko 2.7.** Kontingenssitaulukko lääkäreiden lausunnoista, kun sairaus on harvinainen.

	Terve	Sairas	
Terve	0.93	0.03	$\hat{\pi}_{1.} = 0.96$
Sairas	0.03	0.01	$\hat{\pi}_{2.} = 0.04$
	$\hat{\pi}_{.1} = 0.96$	$\hat{\pi}_{.2} = 0.04$	

Ensimmäisessä tilanteessa havaittu Cohenin kappa  $\hat{\kappa}_C \approx 0.880$ , joka viittaa vahvaan yksimielisyyteen. Toisessa tilanteessa taas yksimielisyys jää heikoksi estimoidun Cohenin kappan perusteella ( $\hat{\kappa}_C \approx 0.219$ ), vaikka havaittu yksimielisyys  $\hat{\theta}_0 = 0.94$  kummassakin tapauksessa. Kuinka tämä ero havaittujen Cohenin kappojen välillä voidaan selittää?

Mielekästä selitystä heikolle yksimielisyydelle taulukon 2.7 tapauksessa on vaikea löytää, mutta se perustuu sattumayksimielisyyden suuruuteen. Nimitään molemmat lääkärit luokittelevat potilaan sairaaksi 96 % todennäköi-

syydellä ja terveeksi 4 % todennäköisyydellä (johtopäätelmät ovat analogisia myös tilanteessa, jossa marginaalisummat arvioijilla ovat erisuuret). Täten, jos lääkärit antaisivat lausuntonsa arvaamalla, todennäköisyydet yhteisille diagnooseille ovat seuraavat:  $P(\text{Kumpikin lääkäri diagnosoi potilaan terveeksi}) = 0.96 \times 0.96 = 0.9216$  ja  $P(\text{Kumpikin lääkäri diagnosoi potilaan sairaaksi}) = 0.04 \times 0.04 = 0.0016$ . Näin ollen sattumayksimielisyys on  $\hat{\theta}_E = 0.9232$ , joka on todella lähellä havaittua yksimielisyyttä  $\hat{\theta}_0 = 0.94$ . Toisin sanoen lähestulkoon koko havaittu yksimielisyys selittyy sattumalla, joten Cohenin kappan arvo jää heikoksi.

Samalla periaatteella ensimmäisessä tilanteessa sattumayksimielisyys  $\hat{\theta}_E \approx 0.500$ , joten päinvastoin kuin toisessa tapauksessa, havaitusta Cohenin kappan arvosta muodostuu melko suuri. Huomataan myös, että ensimmäisessä tilanteessa esiintyvyys on 0.02 ja toisessa tilanteessa 0.92, joka vahvistaa päätelmän, että esiintyvyyden ollessa suuri Cohenin kappa saa pienen arvon suhteessa havaitun yksimielisyyden tasoon. Odotetusti myös hajonta on toisessa tilanteessa suurempi kuin ensimmäisessä tilanteessa ( $0.208 > 0.048$ ).

Sattumalla voidaan selittää, miksi Cohenin kappa saa pienen arvon, jos esiintyvyys on suuri. Kysymys kuuluukin, vaikuttaako sattumayksimielisyys kappan arvoon liikaa suurilla esiintyvyyden tasoilla. Esimerkiksi tilanteissa, joissa ainakin toinen arvioijista on ammattitaitoinen, ei yksimielisyys ole todennäköisesti pelkkää sattumaa. Näin ollen Cohenin kappa aliarvioi todellisen yksimielisyyden tason.

Harha (2.4) vaikuttaa Cohenin kappaan päinvastoin kuin esiintyvyys. Toisin sanoen kun harha on suuri eli erotus  $|\hat{\pi}_{12} - \hat{\pi}_{21}|$  on suuri, saa Cohenin kappa suuren arvon ja päinvastoin. Havainnollistetaan harhan vaikutusta Cohenin kappaan esimerkin avulla.

**Esimerkki 2.3.** Tarkastellaan kahta kuvitteellista tilannetta, jossa kaksi viranomaista tutkii 100 apurahahakemusta, joiden perusteella he päättävät myöntääkö hakijalle apurahaa vai ei. Viranomaiset tekevät päätöksensä toisistaan riippumattomasti, mutta heidän arvionsa perustuvat samoille kriteereille.

Ensimmäisessä tapauksessa viranomaisten välinen havaittu yksimielisyys on todella alhainen  $\hat{\theta}_0 = 0.06$  (taulukko 2.8). Estimoitu sattumayksimielisyys  $\hat{\theta}_E \approx 0.500$ , joten havaittu Cohenin kappa  $\hat{\kappa}_C$  saa arvon  $-0.879$ . Näin ollen viranomaisten välillä vallitsee erimielisyys apurahapäätöksiin suhteen. Havaitun Cohenin kappan arvon ja kontingenssitaulukon perusteella erimielisyys on vahvaa. Tämä viittaa siihen, että luokittelukriteereissä on todennäköisesti joitain kohtia, jotka saattavat aiheuttaa ristiriitaisen tulkinnan ja siten myös eriävät päätökset. Täten havaintojen avulla on mielenkiintoista tutkia syitä erimielisyyteen ja näin parantaa luokittelukriteereitä.

Toisessa tilanteessa  $\hat{\pi}_{12} = 0.93$  ja  $\hat{\pi}_{21} = 0.01$  (taulukko 2.9). Toisin sanoen diagonaalialkiot pysyvät samoina ( $\hat{\pi}_{11} = \hat{\pi}_{22} = 0.03$ ) kuten myös ei-diagonaalialkioiden summa, mutta summan rakenne muuttuu. Nyt estimoitu sattumayksimielisyys  $\hat{\theta}_E \approx 0.077$ , joten havaittu Cohenin kappa  $\hat{\kappa}_C \approx -0.018$ .



**Taulukko 2.8.** Kontingenssitaulukko viranomaisien arvioista, kun pää-  
tösvaihtoehtojen jakauma on melko symmetrinen.

	Kyllä	Ei	
Kyllä	0.03	0.48	$\hat{\pi}_{1.} = 0.51$
Ei	0.46	0.03	$\hat{\pi}_{2.} = 0.49$
	$\hat{\pi}_{.1} = 0.49$	$\hat{\pi}_{.2} = 0.51$	

**Taulukko 2.9.** Kontingenssitaulukko viranomaisien arvioista, kun pää-  
tösvaihtoehtojen jakauma on vino.

	Kyllä	Ei	
Kyllä	0.03	0.93	$\hat{\pi}_{1.} = 0.96$
Ei	0.01	0.03	$\hat{\pi}_{2.} = 0.04$
	$\hat{\pi}_{.1} = 0.04$	$\hat{\pi}_{.2} = 0.96$	

Vaikka kontingenssitaulukoiden perusteella tilanteet muistuttavat toisiaan, on havaittujen Cohenin kappojen ero suuri. Ensimmäisessä tilanteessa harha on 0.02 ja  $\hat{\kappa}_C \approx -0.879$  ( $\hat{\sigma} \approx 0.048$ ). Toisessa tapauksessa harha on 0.92 ja  $\hat{\kappa}_C \approx -0.018$  ( $\hat{\sigma} \approx 0.020$ ). Huomataan, että harhan ollessa suuri, on kapän arvo suurempi kuin tilanteessa, jossa harha saa pienemmän arvon. Tämä ominaisuus säilyy myös tilanteissa, jolloin Cohenin kappaa saa positiivisen arvon. On kuitenkin huomattava, että tapauksissa, joissa  $\kappa_C > 0$ , vaikuttaa harha Cohenin kappaan vähemmän. Tämä johtuu siitä, että Cohenin kapän ollessa positiivinen, tulee myös erotuksen  $\hat{\theta}_0 - \hat{\theta}_E$  olla positiivinen, joten ei-diagonaalialkioiden erotus ei voi olla kovin suuri. Näin ollen harhan vaikutus kapän tulkintaan tulee ilmi nimenomaan silloin, kun Cohenin kappaa on negatiivinen, joka on useimmiten melko harvinaista.

Yllä esitettyjen esimerkkien perusteella on selvää, että on tilanteita, jolloin Cohenin kappaa ei toimi johdonmukaisesti, mistä syystä sen tulkinta hankaloituu. Näin ollen sille on esitetty useita erilaisia muunnoksia ja parannusehdotuksia, jotta sen tulkinta helpottuisi (Sim et al. 2005; Sinha, Pornpis & Montip 2006). Joitain näistä muunnoksista tutkitaan kriittisesti seuraavassa luvussa.

## 3 Cohenin kappan muunnoksia

Cohenin kappa mittaa yksimielisyyttä hyvin useimmissa tilanteissa, mutta on myös tapauksia, joissa se käyttäytyy epäjohdonmukaisesti. Tällaisia tilanteita syntyy muun muassa erimielisyyttä mitattaessa tai esiintyvyyden ollessa todella suuri. Tästä syystä on mielenkiintoista tarkastella erilaisia parannusehdotuksia ja pohtia, mittaavatko ne yksimielisyyttä paremmin. Tässä luvussa esitellään muutamia muunnoksia, joiden tarkoituksena on muuttaa Cohenin kappa siten, että sen tulkinta olisi johdonmukainen myös erikoistilanteissa.

### 3.1 Modifioitu kappa

Lauseen 2.2 (s. 13) yhteydessä perusteltiin, miksi Cohenin kappa ei käyttäydy johdonmukaisesti täydellisen erimielisyyden tilanteessa. Tästä syystä on mielenkiintoista pyrkiä muuntamaan Cohenin kappa siten, että se säilyttäisi omat hyvät ominaisuutensa mutta saisi arvon  $-1$  aina, kun kontingenssitaulukon ei-diagonaalialkioiden summa on 1 ( $\pi_{12} + \pi_{21} = 1$ ).

Muunnetaan Cohenin kappa seuraavasti

$$(3.1) \quad \kappa_{M1} = \frac{\theta_0 - \theta_E}{A - \theta_E}$$

ja ratkaistaan  $A$  ehtojen  $\pi_{11} = \pi_{22} = 0$ ,  $\pi_{12} = \alpha$ ,  $\pi_{21} = 1 - \alpha$ ,  $0 < \alpha < 1$  ja  $\kappa_{M1} = -1$  täytyessä (tilanne vastaa taulukkoa 2.5 s. 13). Täten modifioitu kappa voidaan kirjoittaa muodossa

$$\kappa_{M1} = \frac{-2\alpha(1 - \alpha)}{A - 2\alpha(1 - \alpha)} = -1,$$

josta saadaan, että

$$A = 4\alpha(1 - \alpha).$$

Edelleen ehdoista seuraa, että  $\pi_{1\cdot} = \pi_{2\cdot} = \alpha$ , joten  $\alpha$  voidaan määrittellä seuraavasti

$$\alpha = \frac{\pi_{1\cdot} + \pi_{2\cdot}}{2}.$$

Samalla idealla saadaan, että

$$1 - \alpha = \frac{\pi_{1\cdot} + \pi_{2\cdot}}{2}.$$

Sijoittamalla nämä  $A$ :n lausekkeeseen, voidaan  $A$  määritellä seuraavasti

$$A = 4 \times \frac{\pi_{1.} + \pi_{.2}}{2} \times \frac{\pi_{.1} + \pi_{2.}}{2} = (\pi_{1.} + \pi_{.2})(\pi_{.1} + \pi_{2.}).$$

Näin ollen yhtälön 3.1 perusteella

$$\begin{aligned} \kappa_{M1} &= \frac{\theta_0 - \theta_E}{(\pi_{1.} + \pi_{.2})(\pi_{.1} + \pi_{2.}) - (\pi_{1.}\pi_{.1} + \pi_{.2}\pi_{2.})} \\ &= \frac{\theta_0 - \theta_E}{\pi_{1.}\pi_{.1} + \pi_{1.}\pi_{.2} + \pi_{.2}\pi_{.1} + \pi_{.2}\pi_{2.} - \pi_{1.}\pi_{.1} - \pi_{.2}\pi_{2.}}. \end{aligned}$$

Täten modifioidun kapan määritelmä on

$$(3.2) \quad \kappa_{M1} = \frac{\theta_0 - \theta_E}{\pi_{1.}\pi_{.2} + \pi_{.1}\pi_{2.}},$$

ja se voidaan estimoida lausekkeella

$$(3.3) \quad \hat{\kappa}_{M1} = \frac{\hat{\theta}_0 - \hat{\theta}_E}{\hat{\pi}_{1.}\hat{\pi}_{.2} + \hat{\pi}_{.1}\hat{\pi}_{2.}}.$$

(Sinha et al. 2006.)

Edelleen ollaan kiinnostuneita siitä, toimiiko modifioitu kappia johdonmukaisesti täydellisen yksimielisyyden tilanteessa ja säilyttävätkö se kuitenkin Cohenin kapan hyvät ominaisuudet. Sinha et al. (2006) osoittivat artikkelissaan, että modifioitu kappia käyttäytyy rajatapauksissa toivotulla tavalla. Kyseiset ominaisuudet esitetään lauseissa 3.1– 3.3.

**Lause 3.1.**  $\kappa_{M1} = -1$ , jos ja vain jos  $\pi_{11} = \pi_{22} = 0$ ,  $\pi_{12} = \alpha$  ja  $\pi_{21} = 1 - \alpha$ , kun  $0 < \alpha < 1$ .

**Lause 3.2.**  $\kappa_{M1} = 0$ , jos ja vain jos  $\theta_0 = \theta_E$ .

**Lause 3.3.**  $\kappa_{M1} = 1$ , jos ja vain jos  $\pi_{11} = \beta$ ,  $\pi_{22} = 1 - \beta$ ,  $\pi_{12} = \pi_{21} = 0$ , kun  $0 < \beta < 1$ .

Sen lisäksi, että tutkitaan, mittaako modifioitu kappia samaa asiaa kuin alkuperäinen tunnusluku, on tärkeää tarkastella, säilyttävätkö kappojen arvot saman suuruusluokan vai yliarvioiko vai aliarvioiko modifioitu kappia todellisen Cohenin kapan arvon. Voidaankin osoittaa, että modifioitu kappia on itseisarvoltaan suurempi tai yhtä suuri kuin Cohenin kappia (lause 3.4). Tulos on mielenkiintoinen, sillä se kertoo, että modifioitu kappia yliarvioi Cohenin kapan arvon.

On hyvin tärkeää tutkia sitä, kuinka paljon kapaat eroavat toisistaan ja vaikuttaako erotukseen suuruuteen esimerkiksi kapan merkki, arvioijien välinen yksimielisyys tai otoskoko. Jos kappojen erot ovat suuria, etenkin kun kappia on positiivinen, ei voida ajatella, että kapaat mittaavat samaa asiaa ainakaan samalla mitta-asteikolla. Jos taas erot ovat suhteellisen pieniä, voidaan ajatella,

että modifikaatio on toiminut hyvin ja voidaan olettaa, että kavat mittaavat samaa asiaa eli arvioijien välistä yksimielisyyttä. Kuitenkin tulee muistaa, että kappojen välillä saa olla eroja, kun sillä kuvataan erimielisyyttä. Alaluvussa 6.2 tutkitaan erotuksen suuruutta tarkemmin simulointikokeiden avulla.

**Lause 3.4.** *Modifioidun kavan nimittäjä on  $(\pi_1. - \pi_{.1})^2$  pienempi kuin Cohenin kavan nimittäjä.  $|\kappa_{M1}| \geq |\kappa_C|$ .*

*Todistus.* Osoitetaan, että modifioidun kavan nimittäjä on aina pienempi tai yhtä suuri kuin Cohenin kavan nimittäjä. Tämä todistus riittää osoittamaan, että modifioitu kava on itseisarvoltaan suurempi tai yhtä suuri kuin Cohenin kava, sillä kappojen osoittajat ovat samat ja osoittaja määrää kappojen merkin (toisin sanoen kavat ovat aina samanmerkkiset).

Koska

$1 = 1 \times 1 = (\pi_1. + \pi_{.2})(\pi_{.1} + \pi_{.2}) = \pi_1.\pi_{.1} + \pi_1.\pi_{.2} + \pi_{.2}.\pi_{.1} + \pi_{.2}.\pi_{.2}$ , voidaan nimittäjien erotus ratkaista seuraavasti:

$$\begin{aligned} & \pi_1.\pi_{.2} + \pi_{.1}\pi_{.2} - (1 - \theta_E) \\ &= \pi_1.\pi_{.2} + \pi_{.1}\pi_{.2} - \pi_1.\pi_{.1} - \pi_1.\pi_{.2} - \pi_{.2}.\pi_{.1} - \pi_{.2}.\pi_{.2} + \pi_1.\pi_{.1} + \pi_{.2}.\pi_{.2} \\ &= \pi_1.\pi_{.2} + \pi_{.1}\pi_{.2} - \pi_1.\pi_{.2} - \pi_{.2}.\pi_{.1} \\ &= \pi_1.(1 - \pi_{.1}) + \pi_{.1}(1 - \pi_{.1}) - \pi_1.(1 - \pi_{.1}) - (1 - \pi_{.1})\pi_{.1} \\ &= \pi_1. - \pi_{.1}^2 + \pi_{.1} - \pi_{.1}^2 - \pi_1. + \pi_1.\pi_{.1} - \pi_{.1} + \pi_1.\pi_{.1} \\ &= -\pi_{.1}^2 - \pi_{.1}^2 + 2\pi_1.\pi_{.1} \\ &= -(\pi_1. - \pi_{.1})^2 \leq 0. \end{aligned}$$

Yhtäsuuruus toteutuu, kun  $\pi_1. = \pi_{.1}$  tai  $\pi_{.2} = \pi_{.2}$  eli  $\pi_{12} = \pi_{21}$ . □

Lauseen 3.4 perusteella voidaan päätellä, että modifioidun kavan nimittäjä on  $(\pi_1. - \pi_{.1})^2 = (\pi_{.2} - \pi_{.2})^2 = (\pi_{12} - \pi_{21})^2$  pienempi kuin Cohenin kavan nimittäjä. Tämä tulos voidaan tulkita siten, että mitä suurempi ei-diagonaalialkioiden ero kontingenssitaulukossa on, sitä enemmän kappojen nimittäjät poikkeavat toisistaan. Toisin sanoen heikon yksimielisyyden tapauksessa kavat eroavat toisistaan enemmän kuin vahvan yksimielisyyden tilanteessa. Nimitäin yksimielisyyden ollessa voimakas on eriäviä mielipiteitä vain vähän, joten kappojen erotus jää pieneksi.

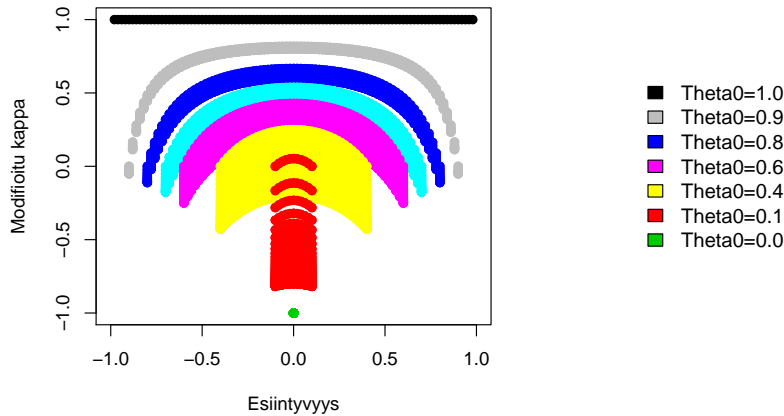
Tämä on hyvä ominaisuus, sillä tavoitteena oli parantaa Cohenin kavan ominaisuuksia nimenomaan erimielisyyden vallitessa. Tulee kuitenkin huomata, että tilanteissa, joissa ei-diagonaalialkiot ovat samaa suuruusluokkaa, on kappojen nimittäjien ero pieni huolimatta siitä, vallitseeko arvioijien välillä voimakas vai heikko yksimielisyys. Toisaalta aikaisemmin esitettiin (s. 16), että Cohenin kava käyttäytyy johdonmukaisesti, kun harha eli ei-diagonaalialkioiden ero on pieni, ja epäjohdonmukaisesti harhan ollessa suuri. Tästä syystä modifioitu kava korjaa Cohenin kappaa oikeaan suuntaan. Havainnollistetaan modifioidun kavan estimointia aikaisemmin esitetyn käytännön esimerkin 2.1 tilanteessa.

**Esimerkki 3.1.** Esimerkin 2.1 tulosten perusteella havaituksi modifioiduksi kapaksi saadaan

$$\hat{\kappa}_{M1} = \frac{\hat{\theta}_0 - \hat{\theta}_E}{\hat{\pi}_1 \cdot \hat{\pi}_2 + \hat{\pi}_1 \cdot \hat{\pi}_2} \approx \frac{\frac{85}{94} - 0.507}{\frac{41}{94} \times \frac{53}{94} + \frac{42}{94} \times \frac{52}{94}} = 0.806$$

Havaituksi Cohenin kapaksi saatiin 0.806 esimerkissä 2.1, joten estimoidut kapat ovat kolmen desimaalin tarkkuudella yhtä suuret. Tulos selittyy sillä, että ero ei-diagonaalialkioiden välillä on pieni  $\frac{5-4}{94} \approx 0.011$  (lause 3.4).

Lauseen 3.4 sivutuotteena osoitettiin, että Cohenin kappa ja modifioitu kappa ovat yhtä suuria, jos ei-diagonaalialkiot ovat samat. Näin ollen lääkäri-esimerkki 2.2 on täysin analoginen modifioidun kapan tilanteessa. Kuvion 3.1 perusteella voidaankin huomata, että esiintyvyys vaikuttaa modifioituun kappaan samoin kuin Cohenin kappaan (vertaa kuvio 2.1, s. 14). Ainoa selkeä ero Cohenin kapan ja modifioidun kapan kuvaajassa on se, että modifioitu kappa saa arvon  $-1$  aina, kun havaittu yksimielisyys  $\theta_0 = 0$ , kun taas Cohenin kapan arvot vaihtelevat välillä  $[-1, 0]$  (lauseet 2.2 ja 3.1).



**Kuvio 3.1.** Modifioidun kapan riippuvuus esiintyvyydestä eri parametrien  $\theta_0$  arvoilla. (Esiintyvyys on laskettu ilman itseisarvoa.)

Harhaesimerkin 2.3 ensimmäisessä tapauksessa havaittu Cohenin kappa ja modifioitu kappa saavat lähestulkoon saman arvon ( $\hat{\kappa}_C \approx -0.879$  ja  $\hat{\kappa}_{M1} \approx -0.880$ ). Toisessa tilanteessa ei-diagonaalialkioiden ero on suurempi, joten kappojen arvot eroavat toisistaan enemmän kuin ensimmäisessä tilanteessa. Yhtälön 3.3 perusteella saadaan, että  $\hat{\kappa}_{M1} \approx -0.219 < \hat{\kappa}_C \approx -0.018$ . Näin ollen harha vaikuttaa modifioituun kappaan, mutta sen vaikutus siihen on heikompi kuin Cohenin kappaan.

Edelleenkin on siis tilanteita, joissa Cohenin kappa ja modifioitu kappa ei käyttäyty johdonmukaisesti tai käyttäytymistä on vaikea selittää. Tästä syystä

Sinha et al. (2006) esittivät artikkelissaan vielä kaksi muunnosta, joita sovelletaan niin Cohenin kappaan kuin myös modifioituun kappaan. Muunnokset perustuvat kappojen standardointiin.

## 3.2 Standardoidut kapat

Standardoinnin ideana on yhteismitallistaa samaa asiaa mittaavien mitta-asteikkojen vaihteluväli. Standardointia hyödynnetään siis tilanteissa, joissa samaa asiaa on mitattu eri mitta-asteikoilla ja tuloksia halutaan verrata keskenään. Toisaalta standardointi Cohenin kappan tai modifioidun kappan yhteydessä saattaa kuulostaa turhalta toimenpiteeltä, ellei niiden vaihteluväliä haluta muuttaa syystä tai toisesta välille  $[0, 1]$ , sillä nehan ovat yhteismitallisia ja vaihtelevat välillä  $[-1, 1]$ . Kuitenkin, jos kappoja tarkastellaan jokin tekijä vakioituna, voidaan huomata, ettei kappojen vaihteluväli olekaan välttämättä  $[-1, 1]$ . Tästä syystä onkin mielenkiintoista tutkia helpottuuko kappojen tulkinta, jos kapat yhteismitallistetaan, jonkin tekijän ollessa vakio.

Sinha et al. (2006) esittivät kaksi standardointia Cohenin kapalle ja modifioidulle kapalle. Toisessa standardointi suoritetaan havaitun yksimielisyyden  $\theta_0$  ja toisessa sattumayksimielisyyden  $\theta_E$  ollessa vakioitu. Kummassakin standardoinnissa idea pohjautuu havainnolle, että vakioinnin seurauksena kappojen maksimi ja minimi vaihtelevat vakioitavan tekijän funktiona. Täten standardointi auttaa yhteismitallistamaan kapat vaihteluvälille  $[0, 1]$ , kun  $\theta_0$  tai  $\theta_E$  on vakioitu. Kuitenkin on syytä pohtia, miten vakiointi vaikuttaa standardoitujen kappojen tulkintaan.

### 3.2.1 Havaittu yksimielisyys vakioitu

Tapauksessa, jossa  $\theta_0$  vakioidaan, tarkastellaan kappan arvoja, kun havaittu yksimielisyys on kiinnitetty tietylle tasolle. Lääkäriesimerkissä 2.2 (s. 15) havaittiin, että vaikka havaittu yksimielisyys pysyi vakiona, havaitun kappan kuvastama yksimielisyyden taso saattoi vaihdella huomattavasti kontingenssitaulukon rakenteesta riippuen. Nämä erot yksimielisyyden tasossa selittyvät sattumayksimielisyyden suuruudella.

Muunnoksia tehtäessä tuleekin pitää mielessä, että kappan määritelmän perusidea on, että sattuman aiheuttama yksimielisyys huomioidaan. Näin ollen on odotettuakin, että jos arvioiden jakauma on kovin vino, toisin sanoen toinen luokka on toista luokkaa huomattavasti harvinaisempi, kasvaa sattumayksimielisyyden suuruus. Tästä syystä luonnollisesti myös kappan arvo pienenee. Mutta varsinkin jos voidaan olettaa, että ainakin toinen arvioijista on ammattitaitoinen ja, että havaittu yksimielisyys on lähellä lukua 1, näyttäisi sattumayksimielisyyden vaikutus kappaan olevan liian suuri jakaumien ollessa vinoja. Näin ollen tutkitaan, helpottaako kappan standardointi sen tulkintaa.

Jotta kapat voidaan suhteuttaa niiden vaihteluvälin suhteen havaitun yksimielisyyden ollessa vakio, tulee määrittää kappojen minimi ja maksimi. Koska Cohenin kappaa ja modifioitu kappaa ovat parametrin  $\theta_E$  laskevia funktioita,

löydetään kummankin kappan ääriarvot sattumayksimielisyyden  $\theta_E$  ääriarvoja tutkimalla. Toisin sanoen maksimoimalla sattumayksimielisyys  $\theta_E$  saadaan ratkaistua kappojen minimi ja päinvastaisesti minimoimalla  $\theta_E$  saadaan selvitettyä kappojen maksimit. Voidaan osoittaa (Sinha et al. 2006), että havaitun yksimielisyyden  $\theta_0$  ollessa vakio, sattumayksimielisyyden  $\theta_E$  ääriarvot ovat seuraavat

$$\theta_{E,min} = \theta_0 \left(1 - \frac{\theta_0}{2}\right) \text{ ja}$$

$$\theta_{E,max} = \theta_0 + \frac{(1 - \theta_0)^2}{2}.$$

Sijoittamalla nämä kappojen yhtälöihin 2.1 ja 3.2 ja sieventämällä saadaan ratkaistua kappojen minimi ja maksimit, jotka ovat

$$(3.4) \quad \kappa_{C,min} = -\left(\frac{1 - \theta_0}{1 + \theta_0}\right), \quad \kappa_{C,max} = \frac{\theta_0^2}{1 + (1 - \theta_0)^2},$$

$$(3.5) \quad \kappa_{M1,min} = -\left(\frac{1 - \theta_0}{1 + \theta_0}\right) \quad \text{ja} \quad \kappa_{M1,max} = \frac{\theta_0}{2 - \theta_0}$$

(Sinha et al. 2006). Modifioidun kappan maksimin määritelmä ei päde, jos  $\theta_0 = 0$ . Nimittäin lauseen 3.1 mukaan modifioitu kappo on  $-1$  aina, jos  $\theta_0 = 0$ , joten maksimi ei voi olla  $0$ .

Määritelmistä huomataan, että kappojen minimi sievenevät samaan muotoon, kun taas maksimit ovat erisuuret. Voidaan osoittaa, että modifioidun kappan maksimi on aina suurempi tai yhtä suuri kuin Cohenin kappan maksimi, kun havaittu yksimielisyys  $\theta_0$  on vakioitu. Näin ollen modifioidun kappan vaihteluväli on aina suurempi tai yhtä suuri kuin Cohenin kappan vaihteluväli.

Nyt koossa on kaikki tarvittava tieto, jotta standardointi parametrin  $\theta_0$  ollessa vakio voidaan suorittaa. Standardoitujen kappojen määritelmät ovat seuraavat:

$$\begin{aligned} \kappa_{M2}(\theta_0) &= \frac{\kappa_{C,max}(\theta_0) - \kappa_C}{\kappa_{C,max}(\theta_0) - \kappa_{C,min}(\theta_0)} \\ &= \frac{\frac{\theta_0^2}{1+(1-\theta_0)^2} - \kappa_C}{\frac{\theta_0^2}{1+(1-\theta_0)^2} + \frac{1-\theta_0}{1+\theta_0}}, \end{aligned}$$

$$\begin{aligned} \kappa_{M3}(\theta_0) &= \frac{\kappa_{M1,max}(\theta_0) - \kappa_{M1}}{\kappa_{M1,max}(\theta_0) - \kappa_{M1,min}(\theta_0)} \\ &= \frac{\frac{\theta_0}{2-\theta_0} - \kappa_{M1}}{\frac{\theta_0}{2-\theta_0} + \frac{1-\theta_0}{1+\theta_0}}. \end{aligned}$$

Koska modifioitu kappo  $\kappa_{M1} = -1$  aina kun havaittu yksimielisyys  $\theta_0 = 0$  (lause 3.1), niin standardoitua kappaa  $\kappa_{M3}(\theta_0)$  ei ole silloin määritelty.

Tarkasteltaessa standardoituja kappoja hiukan tarkemmin huomataan, että niiden tapauksessa pieni arvo viittaa suureen alkuperäiseen kappaan ja suuri arvo pieneen alkuperäiseen kappaan, kun parametri  $\theta_0$  on vakioitu. Koska oletetaan, että parametri  $\theta_0$  on kiinnitetty tietylle tasolle, standardoitujen kappojen nimittäjä pysyy vakiona. Tutkitaan siten kahta tilannetta, joista ensimmäisessä vallitsee vahva yksimielisyys ja päinvastaisesti toisessa tilanteessa vahva erimielisyys. Näin ollen ensimmäisessä tilanteessa alkuperäisen kapan arvo on suuri, ja siten erotuksesta  $\kappa_{max}(\theta_0) - \kappa$  syntyy pieni luku, joten standardoidun kapan arvostakin tulee pieni. Toisessa tilanteessa havaitun kapan arvo on pieni, joten vastaava erotus on suuri ja standardoidun kapan arvo on suuri. Täten standardoitujen kappojen mitta-asteikossa pieni arvo viittaa suureen alkuperäiseen kappaan ja suuri arvo pieneen alkuperäiseen kappaan tietyllä havaitun yksimielisyyden  $\theta_0$  tasolla.

Tulkinnan kannalta voisi olla mielekkäämpää, jos mitta-asteikko olisi päinvastainen eli suuri standardoidun kapan arvo kuvaa suurta alkuperäistä kappa tietyllä havaitun yksimielisyyden tasolla ja toisinpäin. Mitta-asteikko on helppo kääntää päinvastaiseksi, sillä standardoitujen kappojen vaihteluväli on  $[0, 1]$ . Jatkossa tässä tutkimuksessa hyödynnetään Sinhan et al. (2006) määritelmien sijaan seuraavia määritelmiä standardoiduille kapoille, kun havaittu yksimielisyys  $\theta_0$  on vakioitu:

$$(3.6) \quad \kappa_{M2}(\theta_0) = 1 - \frac{\frac{\theta_0^2}{1+(1-\theta_0)^2} - \kappa_C}{\frac{\theta_0^2}{1+(1-\theta_0)^2} + \frac{1-\theta_0}{1+\theta_0}},$$

$$(3.7) \quad \kappa_{M3}(\theta_0) = 1 - \frac{\frac{\theta_0}{2-\theta_0} - \kappa_{M1}}{\frac{\theta_0}{2-\theta_0} + \frac{1-\theta_0}{1+\theta_0}}.$$

Nyt on saatu määritelmä standardoidulle kapalle, kun  $\theta_0$  on vakioitu. Seuraava askel on pohtia, muuttaako vakiointi alkuperäisien kappojen tulkintaa vai voidaanko standardoidulla kapalla korvata alkuperäiset tunnusluvut, toisin sanoen mittaako standardoitu kappo edelleen yksimielisyyttä. Havainnollistetaan standardoinnin ideaa ja siten myös sen tulkintaa kahden esimerkitapauksen avulla.

Oletetaan, että havaittu yksimielisyys  $\theta_0$  on vakioitu nollassa, jolloin Coheinin kappo vaihtelee välillä  $[-1, 0]$  (3.4). Toisaalta, jos havaittu yksimielisyys  $\theta_0$  on suuri, esimerkiksi 0.9, kapan minimi on  $-0.053$  ja maksimi 0.802, joten tässä tilanteessa vaihteluväli on pienempi ja kapan suuruusluokka aivan eri taso. Näin ollen näille kahdelle vaihteluvälille sisältyvät myös erilaiset yksimielisyyden tasot siten, että ensimmäisessä tilanteessa väli kuvaa yksimielisyyttä erimielisyydestä sattumanvaraiseen yksimielisyyteen ja toinen väli sattumanvaraisuudesta vahvaan yksimielisyyteen.

Näin ollen, kun havaittu yksimielisyys vakioidaan, muuttuu tarkasteltavan tilanteen näkökulma täysin. Standardoitua kappa ei voida siis suoraan käyttää yksimielisyyden mittarina. Tarkasti ottaen standardoitu kappo kertoo, kuinka



suuri havaittu kappaa on, muihin kappoihin verrattuna, kun havaittu yksimieliisyys on vakioitu. Havainnollistetaan tulkintaa laskemalla standardoidut kapat edellä esitetyissä tilanteissa.

Oletetaan, että havaittu Cohenin kappaa on  $-0.1$ , kun  $\hat{\theta}_0 = 0$ , joten estimoitu standardoitu kappaa  $\hat{\kappa}_{M2} = 0.9$ . Standardoidun kapan arvo kertoo, että kun havaittu yksimieliisyys  $\hat{\theta}_0 = 0$ , niin siihen suhteutettuna havaittu Cohenin kappaa on todella suuri. Tulee kuitenkin huomata, että yksimieliisyys on vain sattumanvaraista, sillä estimoitu Cohenin kappaa on  $-0.1$ . Näin ollen standardoitu kappaa ei kuvaa millään tavalla yksimieliisyyttä.

Toisessa tapauksessa estimoitu Cohenin kappaa on  $0.7$  ja  $\hat{\theta}_0 = 0.9$ . Nyt havaittu standardoitu kappaa  $\hat{\kappa}_{M2} = 0.881$ , joka kertoo kuinka suuri havaittu Cohenin kappaa on suhteessa muihin kappoihin, kun havaittu yksimieliisyys on  $0.9$ . Siten yksimieliisyyden taso on todella korkea kaikilla tähän vertailuun tulevilla Cohenin kapoilla. Näin ollen vaikka ensimmäisessä tapauksessa havaitun standardoidun kapan arvo on suurempi kuin toisessa tilanteessa, ei se missään nimessä tarkoita, että yksimieliisyys olisi suurempaa siinä tapauksessa. Näiden havaintojen pohjalta voidaan varmistua, ettei standardoitua kappaa voida käyttää yksimieliisyyden mittarina.

Toisaalta standardoidun kapan arvon voi laskea havaitun Cohenin kapan tai modifioidun kapan rinnalle, jotta voidaan arvioida kuinka suuri havaittu kappaa on muihin kappoihin verrattuna, kun havaittu yksimieliisyys on kiinnitetty tietylle tasolle. Tätä tietoa voidaan siis käyttää tulkinnan tukena, mutta alkuperäiset kapat sisältävät olennaisen informaation yksimieliisyyden vahvuuden tulkintaa ajatellen. Havainnollistetaan standardoitujen kappojen, kun havaittu yksimieliisyys  $\theta_0$  on vakio, estimointia ja tulkintaa käytännön aineistolla.

**Esimerkki 3.2.** Standardoidut kapat voidaan estimoida esimerkeissä 2.1 ja 3.1 esitettyjen tulosten avulla seuraavasti:

$$\hat{\kappa}_{M2}(\hat{\theta}_0) = 1 - \frac{\frac{\hat{\theta}_0^2}{1+(1-\hat{\theta}_0)^2} - \hat{\kappa}_C}{\frac{\hat{\theta}_0^2}{1+(1-\hat{\theta}_0)^2} + \frac{1-\hat{\theta}_0}{1+\hat{\theta}_0}} \approx 1 - \frac{\frac{0.904^2}{1+(1-0.904)^2} - 0.806}{\frac{0.904^2}{1+(1-0.904)^2} + \frac{1-0.904}{1+0.904}} = 0.996 \text{ ja}$$

$$\hat{\kappa}_{M3}(\hat{\theta}_0) = 1 - \frac{\frac{\hat{\theta}_0}{2-\hat{\theta}_0} - \hat{\kappa}_{M1}}{\frac{\hat{\theta}_0}{2-\hat{\theta}_0} + \frac{1-\hat{\theta}_0}{1+\hat{\theta}_0}} \approx 1 - \frac{\frac{0.904}{2-0.904} - 0.806}{\frac{0.904}{2-0.904} + \frac{1-0.904}{1+0.904}} = 0.979$$

Huomataan, että  $\hat{\kappa}_{M2}(\hat{\theta}_0)$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_0)$  eroavat hiukan toisistaan, mutta kummatkin ovat todella lähellä ykköstä. Näin ollen niin havaittu Cohenin kappaa kuin myös modifioitu kappaa on lähellä maksimaalista kapan arvoa havaitun yksimieliisyyden ollessa  $0.904$ .

### 3.2.2 Sattumayksimieliisyys vakioitu

Tutkitaan seuraavaksi tilannetta, jossa sattumayksimieliisyys  $\theta_E$  vakioidaan. Tässä tapauksessa kappojen maksimia ja minimiä määriteltäessä tulee tarkas-

telut tehdä kolmessa osassa, jotka ovat  $\theta_E = 0.5$ ,  $\theta_E < 0.5$  ja  $\theta_E > 0.5$ . Näissä tilanteissa ratkaistaan  $\theta_{0,min}$  ja  $\theta_{0,max}$ , kun sattumayksimielisyys  $\theta_E$  on vakioitu. Tämän jälkeen voidaan määrittää ääriarvot Cohenin kapalle ja modifioidulle kapalle. Näin saadaan määritelmät standardoiduille kapoille, kun  $\theta_E$  on vakioitu. (Sinha et al. 2006.)

$$\kappa_{M2}(\theta_E) = \begin{cases} \frac{1-\kappa_C}{2} & \text{kun } \theta_E = 0.5 \\ \frac{1-\kappa_C}{1-\frac{\sqrt{2\theta_E-1}-\theta_E}{1-\theta_E}} & \text{kun } \theta_E > 0.5 \\ \frac{1-\sqrt{1-2\theta_E}-\theta_E-\kappa_C}{\frac{1-\theta_E}{1-\sqrt{1-2\theta_E}}} & \text{kun } \theta_E < 0.5 \end{cases}$$

$$\kappa_{M3}(\theta_E) = \begin{cases} \frac{1-\kappa_{M1}}{2} & \text{kun } \theta_E = 0.5 \\ \frac{(1-\theta_E)(1-\kappa_{M1})}{1-\sqrt{2\theta_E-1}} & \text{kun } \theta_E > 0.5 \\ \frac{1-\sqrt{1-2\theta_E}-\theta_E-\kappa_{M1}}{\frac{\theta_E}{1-\sqrt{1-2\theta_E}}} & \text{kun } \theta_E < 0.5 \end{cases}$$

Jälleen muutetaan tätä määritelmää siten, että mitta-asteikko on käänteinen. Toisin sanoen pieni standardoidun kapan arvo viittaa pieneen alkuperäisen kapan arvoon, kun sattumayksimielisyys on vakioitu ja päinvastoin. Näin ollen standardoitujen kappojen määritelmät, kun parametri  $\theta_E$  on vakioitu, ovat seuraavat:

$$(3.8) \quad \kappa_{M2}(\theta_E) = \begin{cases} 1 - \frac{1-\kappa_C}{2} & \text{kun } \theta_E = 0.5 \\ 1 - \frac{1-\kappa_C}{1-\frac{\sqrt{2\theta_E-1}-\theta_E}{1-\theta_E}} & \text{kun } \theta_E > 0.5 \\ 1 - \frac{1-\sqrt{1-2\theta_E}-\theta_E-\kappa_C}{\frac{1-\theta_E}{1-\sqrt{1-2\theta_E}}} & \text{kun } \theta_E < 0.5 \end{cases}$$

$$(3.9) \quad \kappa_{M3}(\theta_E) = \begin{cases} 1 - \frac{1-\kappa_{M1}}{2} & \text{kun } \theta_E = 0.5 \\ 1 - \frac{(1-\theta_E)(1-\kappa_{M1})}{1-\sqrt{2\theta_E-1}} & \text{kun } \theta_E > 0.5 \\ 1 - \frac{1-\sqrt{1-2\theta_E}-\theta_E-\kappa_{M1}}{\frac{\theta_E}{1-\sqrt{1-2\theta_E}}} & \text{kun } \theta_E < 0.5 \end{cases}$$

Tutkitaan vielä tarkemmin standardoidun kapan tulkintaa sattumayksimielisyyden ollessa vakio. Jos vakioinnin vaikutuksesta standardoinnissa, huomioidaan vain tietyille välille kuuluvat populaation yksimielisyyden tasot, ei standardoitu kappa mittaa enää yksimielisyyttä. Näin ollen, jos voidaan osoittaa,

että mikä tahansa vakioitu sattumayksimielisyyden taso, kuvaa kaikkia populaation yksimielisyyden tasoja, voidaan standardointia käyttäen suoraan yksimielisyyden mittana.

Pohdittaessa miten asiasta voidaan varmistua, tulee muistaa, ettei tarkasteltavan kappan pidä vaihdella välillä  $[-1, 1]$ . Nimittäin standardoinnin tavoitteenahan on korjata alkuperäistä kappa, kun se käyttäytyy epäjohdonmukaisesti. Näin ollen on mahdollista, että yksimielisyyden taso on korkea, mutta kappa saa pienen arvon. Toisin sanoen vaikka kappan arvot eivät vaihtelevälillä  $[-1, 1]$ , saattaa se kuvata skaalaa erimielisyydestä yksimielisyyteen populaatiossa. Täten tutkitaan, voidaanko millä tahansa vakioitun sattumayksimielisyyden  $\theta_E$  arvolla olettaa, että alkuperäisen kappan vaihteluväli kuvaa populaation yksimielisyyttä ääripäästä toiseen.

Arvioidaan tilannetta havaitun yksimielisyyden  $\theta_0$  vaihteluvälin avulla. Nimittäin, jos sen vaihteluväli on todella pieni jollain sattumayksimielisyyden  $\theta_E$  arvolla, voidaan päätellä, että vakioinnin seurauksena tarkasteluun valikoituu vain tietynlaiset yksimielisyyden tasot. Eikä näin ollen standardoitu kappan arvo kuvaa yksimielisyyttä, vaan sillä on samantapainen tulkinta kuin standardoinnilla havaitun yksimielisyyden ollessa vakio.

Sinha et al. (2006) osoittivat, että havaitun yksimielisyyden minimi ja maksimi voidaan esittää seuraavasti, kun sattumayksimielisyyden  $\theta_E$  on vakio:

$$(3.10) \quad [\theta_{0:\min}, \theta_{0:\max}] = \begin{cases} [0, 1] & \text{kun } \theta_E = 0.5 \\ [\sqrt{2\theta_E - 1}, 1] & \text{kun } \theta_E > 0.5 \\ [0, 1 - \sqrt{1 - 2\theta_E}] & \text{kun } \theta_E < 0.5 \end{cases}$$

Huomataan, että jos sattumayksimielisyyden  $\theta_E = 0.5$ , tulevat varmasti kaikki yksimielisyyden tasot tarkasteluun. Kuitenkin ongelmia alkuperäisten kappojen tulkinnassa tulee nimenomaan silloin, kun sattumayksimielisyyden taso on todella suuri, sillä tarkasteltavien luokkien jakauma on hyvin vino. Näin ollen tarkastellaan tilannetta, kun sattumayksimielisyyden  $\theta_E = 0.85$ .

Havaitun yksimielisyyden vaihteluväli on  $[0.837, 1]$  (3.10). Näin ollen karkeasti voidaan arvioida, ettei rajatapauksissa yksimielisyyden taso tule kattavasti edustetuksi. Täten standardointia, kun sattumayksimielisyyden  $\theta_E$  on vakioitu, ei voida pitää yksimielisyyden mittarina. Toki halutessaan tutkija voi käyttää sitä alkuperäisen kappan rinnalla, jos hän kokee sen helpottavan yksimielisyyden tason tulkintaa. Seuraavassa esimerkissä estimoidaan standardoidut kappat  $\hat{\kappa}_{M2}(\hat{\theta}_E)$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_E)$  käytännön esimerkin tapauksessa.

**Esimerkki 3.3.** Esimerkkien 2.1 ja 3.1 tulosten perusteella havaittu yksimielisyyden  $\hat{\theta}_0 = 0.904 > 0.5$ , joten tämän esimerkin tapauksessa standardoidut kappat, kun sattumayksimielisyyden  $\theta_E$  on vakioitu, lasketaan seuraavasti:

$$\hat{\kappa}_{M2}(\hat{\theta}_E) = 1 - \frac{1 - \hat{\kappa}_C}{1 - \frac{\sqrt{2\hat{\theta}_E - 1} - \hat{\theta}_E}{1 - \hat{\theta}_E}} \approx 1 - \frac{1 - 0.806}{1 - \frac{\sqrt{2 \times 0.507 - 1} - 0.507}{1 - 0.507}} = 0.892$$

$$\hat{\kappa}_{M3}(\hat{\theta}_E) = 1 - \frac{(1 - \hat{\theta}_E)(1 - \hat{\kappa}_{M1})}{1 - \sqrt{2\hat{\theta}_E - 1}} \approx 1 - \frac{(1 - 0.507)(1 - 0.806)}{1 - \sqrt{2 \times 0.507 - 1}} = 0.892$$

Jos verrataan näitä kappan arvoja  $\hat{\kappa}_{M2}(\hat{\theta}_0) = 0.996$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_0) = 0.979$ , ovat  $\hat{\kappa}_{M2}(\hat{\theta}_E)$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_E)$  huomattavasti pienempiä. Erot selittyvät yksinkertaisesti sillä, että tunnuslukujen tulkinta on erilainen.

### 3.3 Pohdintaa ja vertailua

Edellä on esitetty muutamia yksimielisyyden mittareita. Seuraavaksi vertailaan kaikkia aiemmin esitettyjä kappoja, ja pohditaan, mistä erot kappojen välillä johtuvat. Lisäksi otetaan kantaa siihen, miten standardoituja kappoja tulisi tulkita ja hyödyntää eri tutkimuksissa.

Tutkitaan kappojen eroja ja tulkintaa kuvitteellisten aineistojen avulla. Tarkastellaan seuraavaksi esimerkkiä 2.2, jossa kaksi lääkäriä antoi lausunnot sadasta potilaasta riippumattomasti ja kiinnostuksen kohteena oli lääkäreiden yksimielisyys. Tilannetta tutkittiin kahden eri kontingenssitaulukon tapauksessa, ja mielenkiinnon kohteena oli löytää selitys, miksi kappojen arvot vaihtelivat, vaikka kontingenssitaulukot muistuttivat läheisesti toisiaan. Kontingenssitaulukot näissä tilanteissa on esitetty taulukossa 3.1.

**Taulukko 3.1.** Kontingenssitaulukot lääkäriesimerkin tapauksessa.

	Terve	Sairas		Terve	Sairas
Terve	0.48	0.03	Terve	0.93	0.03
Sairas	0.03	0.46	Sairas	0.03	0.01

Toisessa vastaavanlaisessa esimerkissä kaksi viranomaista teki apurahapäätöksiä toisistaan riippumattomasti. Taulukossa 3.2 esitetään esimerkkiin liittyvät kontingenssitaulukot.

**Taulukko 3.2.** Kontingenssitaulukot apurahaesimerkin tapauksessa.

	Kylä	Ei		Kyllä	Ei
Kyllä	0.03	0.48	Kyllä	0.03	0.93
Ei	0.46	0.03	Ei	0.01	0.03

Taulukossa 3.3 ovat koottuina kaikkien aiemmin esitettyjen kappojen arvot kummankin esimerkin tapauksessa. Kahdessa ensimmäisessä kapassa vaihteluväli on  $[-1, 1]$  ja muissa  $[0, 1]$  siten, että minimi kuvaa pientä ja maksimi suurta arvoa kyseisessä asteikossa. Koska vaihteluvälit eivät ole kaikilla kapoilla samat, eivätkä ne mittaa samaa asiaa, ei suoranaista vertailua kaikkien kappojen välillä voida suorittaa. Kuitenkin näiden kahden esimerkin perusteella voidaan arvioida alustavasti standardoitujen kappojen toimivuutta ja tulkintaa.

**Taulukko 3.3.** Kappojen vertailu esimerkkien tapauksessa.

	Lääkäriesim.		Apurahaesim.	
	Tilanne 1	Tilanne 2	Tilanne 1	Tilanne 2
Cohenin kappaa	0.880	0.219	-0.879	-0.018
Modifioitu kappaa	0.880	0.219	-0.880	-0.219
$K_{M2}(\theta_0)$	0.999	0.274	0.008	0.977
$K_{M3}(\theta_0)$	0.993	0.272	0.007	0.728
$K_{M2}(\theta_E)$	0.939	0.250	0.061	0.750
$K_{M3}(\theta_E)$	0.939	0.250	0.061	0.750

Lääkäriesimerkin ensimmäisessä tilanteessa kaikki kapat saavat todella korkean arvon suhteessa omaan mitta-asteikkoon. Koska ei-diagonaalialkiot ovat yhtä suuria, on havaittu Cohenin kappaa ja modifioitu kappaa samansuuruisia. Niiden suuri arvo viittaa siihen, että arvioijien välillä on vahva yksimielisyys. Tämä kuvaus on varmasti hyvä arvio yksimielisyyden tasosta, sillä vastaavallisissa tilanteissa Cohenin kappaa (Sim et al. 2005) kuten myös modifioitu kappaa toimivat hyvin. Vaikka myös standardoitujen kappojen arvot ovat suuria, ei niiden perusteella voida päätellä, että yksimielisyyden taso olisi hyvä.

Esimerkiksi havaittu standardoitu kappaa  $\hat{\kappa}_{M2}(\hat{\theta}_0)$  on kahden desimaalin tarkkuudella 1, joten estimoitu Cohenin kappaa on lähestulkoon maksimi kapaa arvo, kun havaittu yksimielisyys  $\hat{\theta}_0 = 0.94$ . Estimoidun standardoidun kapaa  $\hat{\kappa}_{M3}(\hat{\theta}_0)$  tulkinta on sama, mutta standardointi on tehty modifioidulle kapalle.

Kun sattumayksimielisyys  $\theta_E$  on vakioitu, ovat standardoitujen kappojen arvot pienempiä kuin havaitun yksimielisyyden ollessa vakio. Tämä selittyy sillä, että havaitun yksimielisyyden  $\theta_0$  ollessa vakio vangitsee tarkasteltava väli vain pienen yksimielisyyden tason vaihtelun. Tähän vaihteluväliin suhteutettuna havaittu kappaa on todella suuri. Toisaalta taas lääkäriesimerkin ensimmäisessä tilanteessa estimoitu sattumayksimielisyys  $\hat{\theta}_E \approx 0.500$ , joten sattumayksimielisyyden ollessa vakio vaihteluväli sisältää ainakin lähestulkoon koko skaalan erimielisyydestä yksimielisyyteen (vaihteluväli 3.10). Näin ollen nimenomaisesti tässä esimerkissä havaitut standardoidut kapat  $\hat{\kappa}_{M2}(\hat{\theta}_E)$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_E)$  kuvaavat läheisesti yksimielisyyden tasoa vaihteluvälillä  $[0, 1]$ .

Toisessa tilanteessa havaitun modifioidun kapaa ja Cohenin kapaa arvot ovat havaittuun yksimielisyyden tasoon  $\theta_0$  nähden todella pieniä. Asian vahvistavat standardoitujen kappojen pienet arvot. Tulee muistaa, että ainakin osittain kappojen pienuus johtuu taustalla olevan jakauman vinoudesta. Nimittäin arvioijien reunajakaumien perusteella näyttäisi siltä, että otoksessa suurin osa potilaista on terveitä ja vain pieni osa sairaita. Tämä vaikuttaa oletetun sattumayksimielisyyden suuruuteen. Täten sattumayksimielisyydestä muodostuu suuri luku, mistä syystä kapasta tulee pieni luku. Toisin sanoen vinojen jakaumien tapauksessa voisi olla toivottua, että sattumayksimielisyys vaikuttaisi

kapan arvoon vähemmän, varsinkin, jos ollaan varmoja ainakin toisen arvioijan ammattitaidosta. Kuitenkaan sattumayksimieliisyyden  $\theta_E$  vaikutusta ei voida missään nimessä poistaa kokonaan.

Tulee myös huomata, että vaikka lääkäriesimerkin ensimmäisessä tilanteessa voitiin  $\hat{\kappa}_{M2}(\hat{\theta}_E)$  ja  $\hat{\kappa}_{M3}(\hat{\theta}_E)$  avulla arvioida suoraan yksimieliisyyden tasoa, ei sama tulkinta toimi tässä esimerkissä. Nimittäin nyt estimoitu sattumayksimieliisyys  $\hat{\theta}_E \approx 0.923$  ja näin ollen havaittu yksimieliisyys  $\hat{\theta}_0$  vaihtelee välillä  $[0.92, 1]$  (3.10), joten ei voida olettaa, että tarkasteluun tulisi kaikki yksimieliisyyden tasot. Näin ollen lääkäriesimerkki osoittaa, ettei standardointi, kun parametri  $\theta_E$  on vakioitu, mittaa suoraan yksimieliisyyttä niissä tilanteissa, joissa alkuperäinen kappaa toimii epäjohdonmukaisesti. Siten sillä ei voida korvata alkuperäistä kappaa.

Apurahaesimerkissä tilanne on päinvastainen. Nimittäin ensimmäisessä tilanteessa Cohenin kappaa ja modifioitu kappaa saavat todella pienen arvon, joten kyse on erimieliisyydestä. Kaikkien standardoitujen kappojen tulkinta on, että kun jokin tekijä on vakioitu, on havaittu kapaa arvo lähellä minimiä suhteessa muihin kappoihin, sillä kaikki standardoidut kappat saavat todella pienen arvon lähellä nollaa.

Toisessa tilanteessa, näyttäisi siltä, että havaitut Cohenin kapaa ja modifioidun kapaa arvot ovat liian suuria kontingenssitaulukkoon ja ensimmäiseen tilanteeseen verrattuina. Tutkitaan standardoitua kappaa  $\kappa_{M3}$ , kun parametri  $\theta_0$  tai  $\theta_E$  on vakioitu. Havaittujen standardoitujen kappojen arvot ovat melko suuria. Näin ollen havaitun modifioidun kapaa arvo on melko suuri suhteessa muihin kappoihin, kun parametri  $\theta_0$  tai  $\theta_E$  on vakioitu. Täten havaitun modifioidun kapaa arvo on keskimääräistä suurempi muihin ehdon täyttäviin kappoihin verrattuna.

Edellä esitettyjen esimerkkien perusteella voidaan päätellä, että standardoiduilla kappoilla ei voida korvata alkuperäisiä kappoja, sillä ne eivät mittaa suoranaisesti yksimieliisyyttä. Niiden avulla voidaan kuitenkin saada viitettä siitä, kuinka suuri havaittu kappaa on suhteessa muihin kappoihin, kun jokin tekijä on vakioitu. Toisaalta vaikka standardointia voidaan käyttää apuna tulkintaa tehtäessä, on kyseenalaista tarvitseeko sen takia laskea standardoidut kappat vai kannattaisiko apuna käyttää kontingenssitaulukkoa, esiintyvyyttä tai hajontaa. Vai löytyykö kenties jokin muu muunnos, joka korvaisi Cohenin kapaa.

Cohenin kapalle onkin esitetty paljon muita vaihtoehtoisia esitystapoja ja muunnoksia edellä esitettyjen lisäksi. Muun muassa Sim et al. (2005) esittivät artikkelissaan, että havaitun kapaa sijaan voitaisiin käyttää tunnuslukua, josta on poistettu harhan ja esiintyvyyden vaikutus kokonaan. Kuitenkin tätä menettelyä on kritisoitu, sillä se pohjautuu hypoteettiseen tilanteeseen eikä siten kuvaa havaittua tilannetta. Lisäksi esiintyvyys ja harha ovat informatiivisia, joten niiden vaikutusta ei pitäisi poistaa. (Hoehler 2000.) Kuitenkin tilanteissa, joissa esiintyvyys on todella suuri eli tapahtumien jakauma on vino, voi kapaa arvosta muodostua liian pieni todelliseen yksimieliisyyden tasoon nähden. Siten näissä tilanteissa voisi olla perusteltua tulkita kappaa kriittisemmin.

Cohenin kapan tulkinnaassa tuleekin huomioida tekijät, jotka vaikuttavat sen epäjohdonmukaiseen käyttäytymiseen. Ensinnäkin aina arvioijien yksimielisyyden riittävyttä arvioitaessa tulee huomioida tutkittavan asian vakavuus. Toiseksi tulee pohtia, voidaanko olettaa, että ainakin toinen arvioijista on ammattitaitoinen. Jos oletusta ei voida tehdä, ei alkuperäistä kapan arvoa kannata korjata edes vinojen jakaumien tapauksessa, sillä silloin se kuvastaa tilannetta hyvin. Jos taas ainakin toinen arvioijista on ammattitaitoinen, tulee tulkinnaassa huomioida arvioiden jakauman rakenne. Nimittäin vinojen jakaumien tapauksessa sattumayksimielisyyden suuruus yliarvioituu ja näin ollen kapan arvo aliarvioituu. Tällöin on perusteltua huomioida tämä ominaisuus tulkintaa tehtäessä.

Lisäksi muun muassa otoskoko vaikuttaa tulosten luotettavuuteen ja siten myös tulkintaan. Tulkintaa tehtäessä ja raportoitaessa kannattaakin otossuureen  $\hat{k}$  ohessa esittää ainakin kontingenssitaulukko ja otoskoko. Toisaalta pelkkä kapan piste-estimaatti ei huomioi estimointiin liittyvää harhaa, joten sen rinnalla kannattaa esittää myös jokin väliestimaatti kuten luottamusväli. Jotta luottamusväli voidaan estimoida, tulee tutkia Cohenin kapan tilastollisia ominaisuuksia, joita esitellään seuraavassa luvussa.

## 4 Tilastollinen päättely yksimielisyyden mittaamisessa

Tässä luvussa esitellään Cohenin kapin ja modifoidun kapin asymptoottinen odotusarvo ja varianssi. Tulokset johdetaan niin sanotulla Delta-menetelmällä, joka perustuu Taylorin sarjoihin. Aluksi esitellään kappaan liittyvää jakaumateoriaa, jonka jälkeen johdetaan kappojen suurten otosten odotusarvo ja varianssi Delta-menetelmää apuna käyttäen.

### 4.1 Binomi- ja multinomijakauma

Binomi- ja multinomijakaumat ovat tilastotieteen keskeisiä diskreetin satunnaismuuttujan todennäköisyys jakaumia. Molemmilla jakaumilla on myös keskeinen rooli Cohenin kapin ja modifoidun kapin tilastollisia ominaisuuksia johdettaessa, joten perehdytään niiden perusominaisuuksiin lyhyesti. Aloitetaan määrittely Bernoullin jakaumasta, joka yleistetään binomijakaumaksi. Edelleen binomijakauma laajennetaan multinomijakaumaksi ja esitellään sen perusominaisuudet.

Oletetaan, että satunnaismuuttuja  $Y \sim Ber(p)$ , jolloin

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

missä  $0 \leq p \leq 1$ . Ajatellaan, että satunnaismuuttuja  $Y$  kuvaa onnistumista niin sanotussa Bernoullin kokeessa, jossa on kaksi tulosvaihtoehtoa. Näin ollen onnistumisen todennäköisyys on  $P(Y = 1) = p$  ja epäonnistumisen todennäköisyys  $P(Y = 0) = 1 - p = q$ . Satunnaismuuttujan  $Y$  odotusarvo ja varianssi ovat seuraavat:

$$E(Y) = p \quad \text{ja} \quad \text{Var}(Y) = pq$$

Kun toistetaan tätä Bernoullin koetta  $n$ -kertaa, saadaan toisistaan riippumattomien satunnaismuuttujien jono  $Y_1, \dots, Y_n$ . Merkitään nyt onnistumisten lukumäärää satunnaismuuttujalla  $X$ . Onnistumisten lukumäärä noudattaa nyt binomijakaumaa parametrein  $n$  ja  $p$ , missä  $n$  on yritysten lukumäärä ja  $p$  onnistumisen todennäköisyys. Lyhyemmin merkitään  $X \sim Bin(n, p)$ . Tällöin

$$(4.1) \quad E(X) = np,$$

$$(4.2) \quad \text{Var}(X) = npq.$$



Satunnaismuuttujan  $X$  todennäköisyysfunktio on

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

(Johnson & Kotz 1969.)

Jos satunnaismuuttujassa  $X$  olisi kahden tulosvaihtoehdon sijaan useita vaihtoehtoja, noudattaisivat tulokset multinomijakaumaa. Toistetaan siis  $n$  kertaa koetta, jolla on  $k$  tulosvaihtoehtoa. Merkitään näitä tuloksia  $1, 2, \dots, k$ . Olkoon  $p_i$   $i$ :nmen tuloksen todennäköisyys, joten vektori  $\mathbf{p} = (p_1, p_2, \dots, p_k)$  sisältää eri tuloksien todennäköisyydet. Todennäköisyydet eivät muutu eri toistoissa. Olkoon  $X_i$  tuloksen  $i$  lukumäärä, kun koetta toistetaan  $n$  kertaa. Merkitään  $\mathbf{X} = (X_1, X_2, \dots, X_k) \sim Mult(n, \mathbf{p})$

Multinomijakauman todennäköisyysfunktio on

$$f(x_1, \dots, x_k) = \binom{n}{x_1 x_2 \dots x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k},$$

missä  $p_1 + \dots + p_k = 1$  ja  $\binom{n}{x_1 x_2 \dots x_k} = \frac{n!}{x_1! x_2! \dots x_k!}$ . Tiedetään myös, että multinomijakaumassa jokaisen  $X_i$  reunajakauma on binomijakauma eli  $X_i \sim Bin(n, p_i)$ ,  $i = 1, 2, \dots, k$  ja, että  $X_i$  ja  $X_j$  yhteisjakauma on niin sanottu trinomijakauma parametrein  $n, p_i$  ja  $p_j$  (Hogg & Tanis 2006). Täten

$$E(X_i) = np_i, \quad \text{Var}(X_i) = np_i(1-p_i), \quad \text{Cov}(X_i, X_j) = -np_i p_j.$$

(Johnson et al. 1969.)

Määritellään seuraavaksi frekvenssitaulukon elementtien jakaumat (taulukko 4.1). Frekvenssitaulukon solut  $n_{ij}$  noudattavat binomijakaumaa, jos tilan-

**Taulukko 4.1.** Frekvenssitaulukon elementit.

Arvioija A	Arvioija B	Todennäköisyys	Frekvenssi
1	1	$\hat{\pi}_{11}$	$n_{11}$
1	2	$\hat{\pi}_{12}$	$n_{12}$
2	1	$\hat{\pi}_{21}$	$n_{21}$
2	2	$\hat{\pi}_{22}$	$n_{22}$

netta tutkitaan kaksiluokkaisena eli luokka  $n_{ij}$  vastaan muut. Täten  $n_{ij} \sim Bin(n, \pi_{ij})$ , missä  $i, j = 1, 2$  ja  $n = \sum_{i=1}^2 \sum_{j=1}^2 n_{ij}$ . Lausekkeiden 4.1 ja 4.2 perusteella saadaan  $E(n_{ij}) = n\pi_{ij}$  ja  $\text{Var}(n_{ij}) = n\pi_{ij}(1-\pi_{ij})$ . Jos tutkitaan kaikkia frekvenssitaulukon soluja  $(n_{11}, n_{12}, n_{21}, n_{22})$  yhtäaikaisesti, on tarkasteltava jakauma multinomijakauma parametrein  $n$  ja  $\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{22}$ .

Cohenin kappa ja modifioitu kappa voidaan esittää  $n_{ij}$ :n funktiona, joten myös niiden odotusarvo ja varianssi voidaan määrittellä tarkasti yllä esitettyjen jakaumaominaisuuksien avulla. Kuitenkin kappojen lausekkeista tulee todella monimutkaisia, joten tarkastelua helpotetaan approksimoimalla odotusarvo ja varianssi. Käytetty approksimaatio perustuu niin sanottuun Delta-menetelmään, joka puolestaan pohjautuu Taylorin sarjoihin.

## 4.2 Taylorin sarjat

Vuonna 1712 Brook Taylor (1685–1731) esitti Taylorin polynomit, joiden avulla voidaan approksimoida derivoituvia funktioita tietyllä välillä. Olkoon funktio  $f$  jatkuva ja kerran derivoituva pisteessä  $a$ , missä piste  $a$  on jokin mielivaltainen reaaliluku. Asetetaan lisäksi  $P_0(x) = f(a)$ . Paras lineaarinen approksimaatio funktiolle lähellä pistettä  $a$  on lineaarinen polynomi

$$P_1(x) = f(a) + f'(a)(x - a).$$

Polynomilla  $P_1$  on sama arvo ja muutosvauhti pisteessä  $a$  eli  $P_1(a) = f(a)$  ja  $P_1'(a) = f'(a)$ . Jos  $f$  on kaksi kertaa derivoituva pisteessä  $a$ , saadaan sille parempi approksimaatio käyttämällä toisenasteen polynomia

$$P_2(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2.$$

Edelleen polynomilla  $P_2$  on sama arvo pisteessä  $a$  kuin funktiolla  $f$ . Lisäksi niiden kaksi ensimmäistä derivaattaa pisteessä  $a$  ovat samoja. Näin voidaan jatkaa ja saadaan määritelmä Taylorin lauseelle.

**Lause 4.1.** *Taylorin teoreema.*

*Olkoon  $f$   $n + 1$  kertaa derivoituva avoimella välillä  $I$ , joka sisältää pisteen  $a$ . Siten kaikilla  $x \in I$ ,*

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n + R_n(x),$$

missä jäännös  $R_n(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x - t)^n dt$

Voidaan osoittaa, että jos  $R_n(x) \rightarrow 0$ , niin sarjalla on esitysmuoto

$$\begin{aligned} f(x) &= f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \dots \\ &= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x - a)^k \end{aligned}$$

Lausekkeen oikeaa puolta kutsutaan Taylorin sarjaksi  $(x - a)$ :ssa. (Salas, Hille & Etgen 2003.) Hyödynnetään tälle havainnolle perustuvaa Delta-menetelmää funktion  $g$  suurten otosten odotusarvon ja varianssin määrittämiseksi.

Arvioidaan funktiota  $g$  sen odotusarvon ympäristössä Taylorin sarjan kahden ensimmäisen termin avulla. Oletetaan, että  $E(X) = \theta$  ja  $\text{Var}(X) = \sigma^2$ . Arvio on melko tarkka, jos termi  $(X - \theta)$  on pieni, sillä silloin Taylorin sarjan loppuosa lähestyy nollaa.

$$g(X) \approx g(\theta) + g'(\theta)(X - \theta)$$

Näin ollen odotusarvon ominaisuuksien perusteella

$$\begin{aligned}
 E(g(X)) &\approx E(g(\theta) + g'(\theta)(X - \theta)) \\
 &= g(\theta) + g'(\theta)(E(X) - \theta) \\
 (4.3) \qquad &= g(\theta)
 \end{aligned}$$

Saatiin siis tulos, jonka mukaan minkä tahansa funktion odotusarvoa voidaan approksimoida, samalla funktiolla, kun  $x$ :n paikalle sijoitetaan sen odotusarvo.

Approksimoidaan seuraavaksi samalla idealla varianssi ja kovarianssi

$$\begin{aligned}
 \text{Var}(g(X)) &\approx \text{Var}(g(\theta) + g'(\theta)(X - \theta)) \\
 &= (g'(\theta))^2 \text{Var}(X - \theta) \\
 &= (g'(\theta))^2 \text{Var}(X) \\
 (4.4) \qquad &= (g'(\theta))^2 \sigma^2,
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(g(X), f(X)) &= E(g(X) - g(\theta))(f(X) - f(\theta)) \\
 &\approx E((X - \theta)g'(\theta))((X - \theta)f'(\theta)) \\
 &= g'(\theta)f'(\theta)E((X - \theta)^2) \\
 (4.5) \qquad &= \sigma^2 g'(\theta)f'(\theta).
 \end{aligned}$$

Näiden tulosten perusteella voidaan myös approksimoida kappojen odotusarvot ja varianssit.

### 4.3 Harhattomuus ja suurten otosten varianssi

Lähestulkoon poikkeuksetta tuloksia estimoidaan otoksen avulla. Silloin on tärkeää, että populaation tuntematonta parametria arvioidaan sopivan otossuureen avulla, joka antaa keskimäärin oikeita tuloksia. Toisin sanoen piste-estimaattorin tulisi olla harhaton (Casella et al. 2002). Toisaalta piste-estimointiin sisältyy aina epävarmuutta, joten yleensä sen rinnalla halutaan esittää luottamusväli, joka kuvaa piste-estimaattoriin liittyvää satunnaista vaihtelua. Usein ollaan myös kiinnostuneita testaamisesta, jotta voidaan tutkia, onko havaittu arvo tilastollisesti merkitsevä. Jotta testejä ja luottamusvälejä voidaan tehdä, tulee otoksesta estimoitavalle parametrille määrittää varianssi. Tutkitaan aluksi, ovatko otoksesta lasketut Cohenin kappa ja modifioitu kappa harhattomia, jonka jälkeen määritellään niille suurten otosten varianssit, jotta myös niille voidaan laskea luottamusvälejä ja suorittaa testauksia.

Arvioidaan lausekkeen 4.3 perusteella Cohenin kappan ja modifioidun kappan odotusarvot. Merkitään

$$\begin{aligned}
 X &= n_{11} + n_{22}, & X &\sim \text{Bin}(n, \pi_{11} + \pi_{22}), \\
 Y &= n_{1.}, & Y &\sim \text{Bin}(n, \pi_{1.}), \\
 Z &= n_{.1}, & Z &\sim \text{Bin}(n, \pi_{.1}).
 \end{aligned}$$

Näin ollen  $n_{2.} = n - Y$  ja  $n_{.2} = n - Z$ .

$$\begin{aligned}
E(\hat{\kappa}_C) &= E\left(\frac{\hat{\theta}_0 - \hat{\theta}_E}{1 - \hat{\theta}_E}\right) \\
&= E\left(\frac{\frac{n_{11}+n_{22}}{n} - \frac{n_{1.}n_{.1}}{n^2} - \frac{n_{2.}n_{.2}}{n^2}}{1 - \frac{n_{1.}n_{.1}}{n^2} - \frac{n_{2.}n_{.2}}{n^2}}\right) \\
&= E\left(\frac{nX - YZ - (n - Y)(n - Z)}{n^2 - YZ - (n - Y)(n - Z)}\right) \\
&= E\left(\frac{nX - YZ - (n - Y)(n - Z)}{n(Y + Z) - 2YZ}\right) \\
&\approx \frac{n^2(\pi_{11} + \pi_{22} - \pi_{1.}\pi_{.1} - \pi_{2.}\pi_{.2})}{n^2(2\pi_{11} + \pi_{12} + \pi_{21} - 2\pi_{1.}\pi_{.1})} \\
&= \frac{\pi_{11} + \pi_{22} - \pi_{1.}\pi_{.1} - \pi_{2.}\pi_{.2}}{1 - \pi_{22} + \pi_{11} - \pi_{1.}\pi_{.1} - (1 - \pi_{2.})(1 - \pi_{.2})} \\
&= \frac{\theta_0 - \theta_E}{1 - \theta_E} \\
&= \kappa_C
\end{aligned}$$

On siis näytetty, että  $\hat{\kappa}_C$  on asymptoottisesti harhaton estimaattori Cohenin kapalle  $\kappa_C$ . Samalla idealla voidaan myös näyttää, että  $E(\hat{\kappa}_{M1}) \approx \kappa_{M1}$ , joten  $\hat{\kappa}_{M1}$  on asymptoottisesti modifioidun kapan harhaton estimaattori (Pornpis 2007).

Cohenin kapan varianssille on esitetty erilaisia esitysmuotoja. Muun muassa Everitt (1968) osoitti hypergeometrisen jakaumaan perustuvan tarkan esityksen Cohenin kapan varianssille. Kuitenkin rutiininomaiseen käyttöön määrittelmä on turhan hankala, joten hän esitti myös erään approksimaation sille. Osoittautui kuitenkin, että approksimaatio on harhainen, joten se vaikuttaa muun muassa testauksien tuloksiin. Useimmilla muillakin Cohenin kapan varianssin approksimaatiolla on sama ongelma (Fleiss, Cohen & Everitt 1969). Vuonna 1969 Fleiss et al. esittivät Cohenin kapan varianssille paremman approksimaation, joka on yleisesti käytössä

$$\begin{aligned}
(4.6) \quad Var(\kappa_C) &\approx \frac{1}{N(1 - \theta_E)^4} \left\{ \sum_{i=1}^2 \pi_{ii} [(1 - \theta_E) - (\pi_{.i} + \pi_{i.})(1 - \theta_0)]^2 \right. \\
&\quad \left. + (1 - \theta_0)^2 \sum_{i \neq j} \pi_{ij} (\pi_{.i} + \pi_{j.})^2 - (\theta_0 \theta_E - 2\theta_E + \pi_0)^2 \right\}.
\end{aligned}$$

Määritellään approksimoitu varianssi myös modifioidulle kapalla. Approksimaatio voidaan johtaa Delta-menetelmällä. Lausekkeisiin 4.3–4.5 verrattuna modifioidun kapan approksimoidun varianssin johtaminen, on hiukan hankalampaa, sillä se riippuu yhden parametrin sijaan kolmesta parametrusta. Esite-

tään modifioitu kappa  $X$ ,  $Y$  ja  $Z$ :n funktiona.

$$\begin{aligned}
\hat{\kappa}_{M1} &= \frac{\hat{\theta}_0 - \hat{\theta}_E}{\hat{\pi}_1 \cdot \hat{\pi}_2 + \hat{\pi}_{\cdot 1} + \hat{\pi}_{\cdot 2}} \\
&= \frac{\frac{n_{11} + n_{22}}{n} - \frac{n_{1 \cdot} n_{\cdot 1}}{n^2} - \frac{n_{2 \cdot} n_{\cdot 2}}{n^2}}{\frac{n_{1 \cdot} n_{2 \cdot}}{n^2} - \frac{n_{\cdot 1} n_{\cdot 2}}{n^2}} \\
&= \frac{nX - [YZ + (n - Y)(n - Z)]}{Y(n - Y) + Z(n - Z)} \\
&= \frac{n(X + Y + Z) - n^2 - 2YZ}{n(Y + Z) - (Y^2 + Z^2)} \\
&= \psi(X, Y, Z)
\end{aligned}$$

Nyt lausekkeiden 4.3– 4.5 perusteella

$$\begin{aligned}
\text{Var}[\psi(X, Y, Z)] &= \text{Var}(X) \left( \frac{\partial \psi}{\partial X} \right)^2 + \text{Var}(Y) \left( \frac{\partial \psi}{\partial Y} \right)^2 + \text{Var}(Z) \left( \frac{\partial \psi}{\partial Z} \right)^2 \\
(4.7) \quad &+ 2\text{Cov}(X, Y) \frac{\partial \psi}{\partial X} \frac{\partial \psi}{\partial Y} + 2\text{Cov}(X, Z) \frac{\partial \psi}{\partial X} \frac{\partial \psi}{\partial Z} + 2\text{Cov}(Y, Z) \frac{\partial \psi}{\partial Y} \frac{\partial \psi}{\partial Z},
\end{aligned}$$

missä

$$\begin{aligned}
E(X) &= n(\pi_{11} + \pi_{22}), & E(Y) &= n\pi_{1 \cdot}, & E(Z) &= n\pi_{\cdot 1}, \\
\text{Var}(X) &= n\theta_0(1 - \theta_0), & \text{Var}(Y) &= n\pi_{1 \cdot}(1 - \pi_{1 \cdot}), & \text{Var}(Z) &= n\pi_{\cdot 1}(1 - \pi_{\cdot 1}), \\
\text{Cov}(X, Y) &= -n\theta_0\pi_{1 \cdot}, & \text{Cov}(X, Z) &= -n\theta_0\pi_{\cdot 1}, & \text{Cov}(Y, Z) &= -n\pi_{1 \cdot}\pi_{\cdot 1}
\end{aligned}$$

ja

$$\begin{aligned}
\frac{\partial \psi}{\partial X} &= n, \\
\frac{\partial \psi}{\partial Y} &= \frac{(n - 2Z)[n(Y + Z) - (Y^2 + Z^2)] - [n(X + Y + Z) - n^2 - 2YZ](n - 2Y)}{[n(Y + Z) - (Y^2 + Z^2)]^2}, \\
\frac{\partial \psi}{\partial Z} &= \frac{(n - 2Y)[n(Y + Z) - (Y^2 + Z^2)] - [n(X + Y + Z) - n^2 - 2YZ](n - 2Z)}{[n(Y + Z) - (Y^2 + Z^2)]^2}.
\end{aligned}$$

Derivaattoihin sijoitetaan parametrien odotusarvot, jonka jälkeen tehdään sijoitus lausekkeeseen 4.7.

Pornpis (2007) on näyttänyt väitöskirjassaan, että modifioidun kappan suurten otosten varianssi voidaan kirjoittaa muodossa

$$\begin{aligned}
\text{Var}[\psi(X, Y, Z)] &= \frac{1}{n(\pi_{1 \cdot} \pi_{2 \cdot} + \pi_{\cdot 1} \pi_{\cdot 2})^2} \sum_{i \neq j=1}^2 \pi_{ii} \left\{ (\pi_{j \cdot} - \pi_{\cdot i}) - \kappa_{M1}(2\pi_{jj} + 1 - \theta_0) \right\} \times \\
&\quad \left[ (1 + \kappa_{M1})(\theta_0(1 - \theta_0) + 4\pi_{ij}\pi_{ji}) + (1 - \pi_{ii} + \pi_{jj})(\pi_{j \cdot} - \pi_{\cdot i}) \right. \\
&\quad \left. + \kappa_{M1}(2\pi_{jj}(2\pi_{ii} - 1) - (1 - \theta_0)^2) \right] + (1 + \kappa_{M1})\pi_{ji}(\pi_{i \cdot} + \pi_{\cdot j}) \times \\
&\quad \left[ (1 + \kappa_{M1})(\theta_0^2 + 2\pi_{ij}(1 - 2\pi_{ji})) - (\pi_{j \cdot} - \pi_{\cdot i})^2 - \right. \\
(4.8) \quad &\left. \kappa_{M1}(4\pi_{ii}\pi_{jj} + \theta_0(1 - \theta_0)) \right]
\end{aligned}$$

Nyt on määritelty suurten otosten varianssi Cohenin kapalle ja modifioidulle kapalle. Varianssi huomioi otoskoon ja antaa hyvää viitettä siitä, kuinka tarkka otoksesta laskettu kapan arvo on. Toisaalta tulkintaa tehtäessä ei tarvitse tyytyä pelkkään hajontaan, sillä sen avulla kapalle voidaan määrittää luottamusväli ja suorittaa testauksia.

## 4.4 Testaus

Jotta otantavirhe voidaan huomioida otoksesta lasketussa kapan arvossa, suoritetaan testaus tai lasketaan luottamusväli piste-estimaatin rinnalle. Testauksessa voidaan käyttää yksi- tai kaksisuuntaista testausta. Sim et al. (2005) esittivät, että Cohenin kappa tutkiessa yksisuuntaista testiä olisi mielekäs käyttää tilanteissa, joissa nollahypoteesina on, että kappa on nolla. Perusteluna on se, että negatiivisella kapan arvolla ei ole yleensä tulkintaa. Muissa tapauksissa Sim et al. (2005) kehottavat käyttämään kaksisuuntaista testausta, sillä ei ole teoreettista syytä olettaa, että testitulos olisi välttämättä suurempi kuin asetettu kynnsarvo.

Käytännössä ollaan lähes aina kiinnostuneita tilanteista, joissa arvioijien yksimielisyys on parempaa kuin sattumayksimielisyys. Näin ollen useimmiten nollahypoteesiksi asetettu kynnsarvo on suurempi kuin nolla. Se, kuinka suureksi kynnsarvo asetetaan, riippuu asetettujen luokkien erillisyydestä, tutkitavan tilanteen vakavuudesta ja otosjakaumien ominaisuuksista (esimerkiksi esiintyvyys).

Merkitään kynnsarvoa  $\kappa_0$  ja otoksesta laskettua kappaa  $\hat{\kappa}$ . Mahdolliset hypoteesit ovat seuraavat:

$$\begin{aligned} H_0 : \hat{\kappa} &= \kappa_0 \\ H_1 : \hat{\kappa} &\neq \kappa_0 && \text{(kaksisuuntainen testaus)} \\ H_1 : \hat{\kappa} &> \kappa_0 && \text{(yksisuuntainen testaus)} \\ H_1 : \hat{\kappa} &< \kappa_0 && \text{(yksisuuntainen testaus)} \end{aligned}$$

Suurten otosten teorian perusteella

$$K_z = \frac{\hat{\kappa} - \kappa_0}{\sqrt{\text{Var}(\hat{\kappa})}} \stackrel{\text{lik.}}{\sim} N(0, 1).$$

Testisuureeseen  $K_z$  liittyvä pienin riskitaso, jolla  $H_0$  voidaan hylätä, katsotaan standardoidun normaalijakaumana taulukosta (yksisuuntainen testauksessa p-arvo  $= 1 - \Phi(|K_z|) = 1 - P(Z \leq |K_z|)$  ja kaksisuuntainen testauksessa p-arvo  $= 2 * (1 - \Phi(|K_z|))$ , missä  $Z \sim N(0, 1)$ ), joten  $H_0$  hylätään, jos p-arvo on pienempi kuin asetettu merkitsevyystaso. (Hogg et al. 2006.)

Kuitenkin lopulliset päätelmät riippuvat asetetusta hypoteesista. Jos yksisuuntaisessa testauksessa  $H_1 : \hat{\kappa} > \kappa_0$  ja nollahypoteesi hylätään, tehdään päätelmä, että yksimielisyyden taso on riittävä. Hypoteesin  $H_1 : \hat{\kappa} < \kappa_0$  tapauksessa päätelmä on päinvastainen. Kaksisuuntaisessa testauksessa nollahypoteesin

hylkääminen kertoo ainoastaan sen, eroaako havaittu kappala tilastollisesti merkittävästi kynnysarvosta. Jos siis vaihtoehtoinen hypoteesi hyväksytään, tulee tutkia testisuureen arvon merkki, jotta voidaan tehdä päätelmä, onko yksimielisyys riittävää vai ei. Käytännössä on suositeltavaa käyttää kaksisuuntaista testiä, sillä otoksesta lasketun kappala arvon suuruudesta ei yleensä voida tehdä oletuksia.

Testauksen lisäksi luottamusväli on hyvin informatiivinen. Se kertoo, millä välillä populaation kappala on, kun merkitsevyytasoksi on asetettu  $\alpha$ . Kappala  $100(1 - \alpha)$  %:n luottamusväli voidaan määrittää seuraavasti:

$$\left( \hat{\kappa} - z_{\alpha/2} \sqrt{\text{Var}(\hat{\kappa})}, \hat{\kappa} + z_{\alpha/2} \sqrt{\text{Var}(\hat{\kappa})} \right),$$

missä  $\Phi(z_{\alpha/2}) = P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$ . (Hogg et al. 2006.)

Kynnysarvon sisältyessä luottamusvälille voidaan tehdä päätelmä, että merkitsevyytasolla  $\alpha$  kappala on nollassa hypoteesiksi asetettu  $\kappa_0$ . Toisaalta luottamusvälin alarajan ollessa suurempi kuin asetettu kynnysarvo tehdään päätelmä, että populaation kappala arvo on suurempi kuin  $\kappa_0$  merkitsevyytasolla  $\alpha$ , joten yksimielisyys on riittävää. Jos taas luottamusvälin yläraja on pienempi kuin  $\kappa_0$ , ei yksimielisyys ole riittävän vahvaa.

Luottamusväli on siinä mielessä testausta informatiivisempi tapa tarkastella tilannetta, että kun testauksen avulla voidaan päätellä, onko kappala tilastollisesti merkittävästi suurempi kuin kynnysarvo, nähdään luottamusvälin avulla lisäksi kappala vaihteluväli. Vaihteluvälin avulla voidaan arvioida muun muassa estimoinnin tarkkuutta ja tutkia tilannetta monipuolisemmin.

On hyvä huomata, että jotta luottamusvälin ja testauksen avulla tehdään sama päätelmä, tulee kiinnittää erityistä huomiota kynnysarvon asettamiseen. Hypoteeseja asettaessa tulee miettiä, tulisiko kappala olla tilastollisesti merkittävästi suurempi kuin kynnysarvon  $\kappa_0$  vai sallitaanko yhtäsuuruus. Tämä taas vaikuttaa luottamusväliä tulkittaessa siihen, saako kynnysarvo olla luottamusvälillä vai ei.

Oletetaan, että nollassa hypoteesi on asetettu siten, että kappala tulisi olla suurempaa kuin se, jotta yksimielisyys on riittävää. Tällöin jos havaittu arvo ei ole tilastollisesti merkittävästi suurempi kuin asetettu kynnysarvo, tulisi luokittelun perusteita parantaa tai arvioijia kouluttaa lisää. Näissä tilanteissa kannattaa pohtia tarkasti, missä tapauksissa arvioijat ovat yksimielisiä ja erityisesti yrittää selvittää, onko joitain tiettyjä syitä tai tilanteita, jolloin arvioijat ovat erimielisiä. Isoilla aineistoilla tämä voi olla hankalaa, sillä tarkastelua ei voida tehdä missään nimessä havaintokohtaisesti. Kuitenkin yksi tapa tutkia tilannetta on muodostaa muuttuja, joka kertoo, ovatko arvioijat yksimielisiä vai ei ja tutkia tämän muuttujan riippuvuutta eri taustamuuttujista.

**Esimerkki 4.1.** Suoritetaan testaus ja lasketaan luottamusväli Cohenin kappala käytännön aineiston tapauksessa (vertaa tulokset esimerkissä 2.1). Sama voitaisiin tehdä analogisesti modifoidulle kappala.

Esiintyvyys on 0.117, joten Cohenin kappan pitäisi toimia johdonmukaisesti. Täten sitä ei tarvitse huomioida kynnsarvoa määrittäessä. Toisaalta mielipaikkatutkimuksen luonteen ja luokittelun tarkkuuden perusteella yksimielisyyden tason ei tarvitse olla huippuluokkaa. Näin ollen asetetaan, että yksimielisyys on riittävä, jos voidaan olettaa, että populaation kappa on tilastollisesti merkittävästi suurempi kuin 0.6. Karkeasti arvioituna yksimielisyys olisi tuolloin vähintäänkin huomattavaa (Landis et al. 1977).

Asetetaan merkitsevyystasoksi 5 % ja suoritetaan testaus kaksisuuntaisena, joten

$$H_0 : \hat{\kappa}_C = 0.6$$

$$H_1 : \hat{\kappa}_C \neq 0.6.$$

$\hat{\kappa}_C \approx 0.806$  ja  $\sqrt{\text{Var}(\hat{\kappa}_C)} \approx 0.062$ , joten testisuureen arvo on

$$K_Z = \frac{\hat{\kappa} - \kappa_0}{\sqrt{\text{Var}(\hat{\kappa})}} \approx \frac{0.806 - 0.6}{0.062} \approx 3.347$$

ja  $p = 2 \times (1 - \Phi(3.347)) \approx 0.001$ . Pienin riskitaso, jolla  $H_0$  voidaan hylätä, on pienempi kuin 0.05, joten  $H_0$  hylätään. Näin ollen 5 % merkitsevyystasolla kappa eroaa tilastollisesti merkittävästi luvusta 0.6. Lisäksi sillä  $\hat{\kappa}_C > \kappa_0$ , on populaatiossa yksimielisyys 5 % merkitsevyystasolla vähintään huomattavaa ja siten myös riittävää.

Tarkastellaan tilannetta lisäksi 95 % luottamusvälin avulla.

$$\begin{aligned} & \left( \hat{\kappa}_C - z_{\alpha/2} \sqrt{\text{Var}(\hat{\kappa}_C)}, \hat{\kappa}_C + z_{\alpha/2} \sqrt{\text{Var}(\hat{\kappa}_C)} \right) \\ &= (0.806 - 1.96 \times 0.062, 0.806 + 1.96 \times 0.062) \\ &= (0.685, 0.926) \end{aligned}$$

Luottamusvälin avulla voidaan tehdä sama päätelmä kuin testauksessa, sillä 0.6 ei kuulu luottamusvälille ja luottamusvälin alaraja on suurempi kuin kynnsarvo. On myös informatiivista nähdä vaihteluväli, jolla populaation Cohenin kappa on 5 % merkitsevyystasolla.



## 5 Otoksoon määrittäminen

Cohenin kapan suurten otosten varianssin tunteminen mahdollistaa otoksoon määrittämisen ennen tutkimuksen suorittamista. Näin ollen on mahdollista varmistaa, että testauksessa saavutetaan haluttu voimakkuus tai, että luottamusväli on tarpeeksi kapea. Otoksoon määrittämistä voidaan lähestyä kahdella tavalla. Toisaalta voidaan asettaa testaukselle haluttu voimakkuus tai määrätä luottamusvälille sopiva leveys. Kummassakin tapauksessa otoksoon valinta perustuu Cohenin kapan  $\kappa_C$  maksimaaliseen suurten otosten hajontaan (Flack, Afifi & Lachenbruch 1988). Otokskoko voidaan määrittää myös modifioitua kappa käyttäen, mutta tulokset ovat lähes samat, sillä tarkastelu rajoittuu yksimielisyyteen.

### 5.1 Luottamusväli perusteinen

Yksimielisyytutkimuksessa otokskoko tulisi määrittää siten, että estimoidun Cohenin kapan  $\hat{\kappa}_C$  hajonta ei olisi liian suuri. Fleiss et al. (1969) esittivät Cohenin kapan suurten otosten varianssin lausekkeen (4.6), joka voidaan kirjoittaa muodossa

$$\begin{aligned} \text{Var}(\kappa_C) &\approx \frac{1}{N(1-\theta_E)^4} \left\{ \sum_{i=1}^2 \pi_{ii} [(1-\theta_E) - (\pi_{\cdot i} + \pi_i)(1-\theta_0)]^2 \right. \\ &\quad \left. + (1-\theta_0)^2 \sum_{i \neq j} \pi_{ij} (\pi_{\cdot i} + \pi_j)^2 - (\theta_0 \theta_E - 2\theta_E + \pi_0)^2 \right\} \\ &= \frac{Q}{N}. \end{aligned}$$

Voidaan osoittaa, että Cohenin kapan suurten otosten varianssi on esitettävissä parametrien  $\pi_{\cdot 1}$ ,  $\pi_1$  ja  $\kappa_C$  funktiona. Jos nämä suureet ovat tunnettuja, muut termit voidaan määrittää niiden perusteella. (Cantor 1996.)

Tavoitteena on saada  $100(1-\alpha)\%$  luottamusväli Cohenin kapalle siten, että sen leveys on korkeintaan  $2d$ . Tutkijan tulee arvioida parametrit  $\pi_{\cdot 1}$ ,  $\pi_1$  ja  $\kappa_C$  esimerkiksi aiemman tutkimuksen perusteella. Tämän jälkeen voidaan määrittää Cohenin kapan suurten otosten hajonta  $\sqrt{\text{Var}(\kappa_C)}$ , joten

$$\begin{aligned} Z_{\alpha/2} \sqrt{\text{Var}(\kappa_C)} &\leq d \\ Z_{\alpha/2} \sqrt{\frac{Q}{N}} &\leq d \end{aligned}$$

$$N \geq \frac{z_{\alpha/2}^2 Q}{d^2}$$

Jos  $\pi_{.1}$ ,  $\pi_1$  ja  $\kappa_C$  halutaan määrittää tarkkojen arvojen sijaan jollain välillä tai, jos Cohenin kapin arvosta ei haluta tehdä oletuksia etukäteen, tulee otoskoko määrittäessä käyttää maksimaalista parametrin  $Q$  arvoa. Cantor (1996) esitti artikkelissaan taulukoita parametrin  $Q$  maksimi-arvoista erilaisilla marginaalitodennäköisyyksien ja Cohenin kapin arvoilla (Liitteet A ja B).

**Esimerkki 5.1.** Halutaan määrätä otoskoko 95 % luottamusvälin avulla. Aikaisemman tutkimuksen perusteella voidaan arvioida, että  $\pi_{.1} = 0.4$ ,  $\pi_1 = 0.4$  ja  $\kappa_C = 0.6$ , joten  $Q = 0.668$  (Liite B). Luottamusvälin leveys saa olla 0.2, joten  $d = 0.1$ .

$$N \geq \frac{z_{\alpha/2}^2 Q}{d^2} = \frac{1.96^2 \times 0.668}{0.1^2} \approx 257$$

Jotta luottamusvälin leveys on maksimissaan 0.2, tulee tutkimuksessa otoskoon olla vähintään 257.

Jos samassa tilanteessa ei ole perusteluja arvioida Cohenin kappa, valitaan maksimaalinen parametrin  $Q$  arvo, kun  $\pi_{.1} = 0.4$  ja  $\pi_1 = 0.4$ . Tässä tapauksessa  $Q \approx 1.006$  (Liite A), joten  $N > 386$ .

Esimerkin tapauksessa huomataan, että lisäoletus Cohenin kapin suuruudesta pienentää otoskoko huomattavasti. Otoskoon suuruuteen vaikuttaa myös halutun luottamusvälin kapeus sekä merkitsevyystason ja hajonnan suuruus.

## 5.2 Testaus perusteinen

Oletetaan, että tavoitteena on suorittaa testaus merkitsevyystasolla  $\alpha$ . Merkitsevyystaso tarkoittaa todennäköisyyttä saada nollahypoteesin mukaisesta perusjoukosta otos, jolla nollahypoteesi hylätään, vaikka se olisi tosi. Toisaalta on määriteltävä kuinka suuri poikkeama nollahypoteesin mukaisesta kapasta on niin suuri, että se halutaan testissä havaita. Tätä testin kykyä havaita erilaiset poikkeamat nollahypoteesista kutsutaan testin voimakkuudeksi. Merkitään tätä todennäköisyyttä  $1 - \beta$ . Oletetaan lisäksi, että satunnaisvaihtelun määrää kuvaa kapin asymptoottista hajontaa. (Casella et al. 2002.)

Tarkastellaan siis tilannetta, jossa testataan  $H_0 : \kappa_C = \kappa_0$  merkitsevyystasolla  $\alpha$  ja voimakkuudella  $1 - \beta$ , jos  $\kappa_C = \kappa_1$ . Merkitään, että  $b$  on kriittinen raja ja oletetaan, että testaus suoritetaan yksisuuntaisena. Nyt määritelmien

perusteella saadaan

$$\begin{aligned}\Phi\left(\frac{b - \kappa_1}{\sqrt{\frac{Q_1}{N}}}\right) &= \beta \\ \frac{\kappa_0 + Z_\alpha\sqrt{\frac{Q_0}{N}} - \kappa_1}{\sqrt{\frac{Q_1}{N}}} &= -Z_\beta \\ \kappa_0 - \kappa_1 + Z_\alpha\sqrt{\frac{Q_0}{N}} &= -Z_\beta\sqrt{\frac{Q_1}{N}} \\ \kappa_0 - \kappa_1 &= -Z_\beta\sqrt{\frac{Q_1}{N}} - Z_\alpha\sqrt{\frac{Q_0}{N}} \\ \sqrt{N} &= \frac{Z_\alpha\sqrt{Q_0} + Z_\beta\sqrt{Q_1}}{\kappa_1 - \kappa_0},\end{aligned}$$

joten

$$N = \left[ \frac{Z_\alpha\sqrt{Q_0} + Z_\beta\sqrt{Q_1}}{\kappa_1 - \kappa_0} \right]^2$$

(Cantor 1996). Otokokoa määrittäessä ei ole mielekästä asettaa kaksisuuntaista testiä, joten sitä ei esitetä tässä yhteydessä. Havainnollistetaan otokoon määrittämistä kuvitteellisen esimerkin avulla.

**Esimerkki 5.2.** Tutkimuksessa tullaan olemaan kiinnostuneita testausasetelmasta, jossa  $H_0 : \kappa_C = 0.4$  ja vaihtoehtoinen hypoteesi  $H_1 : \kappa_C > 0.4$ . Otokoko halutaan määrittää siten, että testauksessa saavutetaan merkitsevyystaso 0.05 ja, että voimakkuudella 0.80 testi hylkää nollahypoteesin Cohenin kapan arvolla 0.6. Nyt  $Z_\alpha = 1.645$  ja  $Z_\beta = 0.842$ . Oletetaan lisäksi, että  $\pi_1 = \pi_0 = 0.5$ , joten  $Q_0 = 0.840$  ja  $Q_1 = 0.640$ . Näin ollen

$$\begin{aligned}N &= \left[ \frac{Z_\alpha\sqrt{Q_0} + Z_\beta\sqrt{Q_1}}{\kappa_1 - \kappa_0} \right]^2 \\ &= \left[ \frac{1.645\sqrt{0.840} + 0.842\sqrt{0.640}}{0.6 - 0.4} \right]^2 \\ &\approx 119\end{aligned}$$

Jotta testaus tuottaa vähintään ennalta määrätyn tarkkuuden, tulee otokseen valita ainakin 119 toisistaan riippumatonta havaintoa.

## 6 Simulointi

Havaitun aineiston avulla ei aina päästä riittävään tarkkuuteen arvioitaessa koko populaatiota koskevien parametrien ominaisuuksia. Simulointi mahdollistaa tarkemman estimaattoreiden ominaisuuksien tutkimisen äärellisellä havaintoaineistolla. Efron (1979) esitti Bootstrap-menetelmän, joka perustuu Jackknife-menetelmään. Bootstrap-menetelmän ideana on tehdä otoksia palauttaen alkuperäisestä havaintoaineistosta. Seuraavissa luvuissa tarkastellaan Cohenin kapian ja modifioidun kapian ominaisuuksia Bootstrap-menetelmää apuna käyttäen.

### 6.1 Bootstrap-menetelmä

#### 6.1.1 Teoria

Olkoon  $y_1, y_2, \dots, y_n$  otos palauttaen riippumattomista ja samoin jakautuneista satunnaismuuttujista  $Y_1, Y_2, \dots, Y_n$ . Merkitään näiden satunnaismuuttujien todennäköisyysfunktioita  $f$  ja kertymäfunktioita  $F$  vastaavasti. Bootstrap-menetelmän ideana on muodostaa empiirinen jakauma  $\hat{F}$ , jossa jokaisella otoshavainnolla on sama todennäköisyys  $n^{-1}$ . (Davison & Hinkley 1997.)

Halutaan estimoida populaation parametria  $\theta$ , jonka estimaattori on  $\hat{\theta}$ . Oletetaan, että  $\hat{\theta}$  on empiirisen jakauman  $\hat{F}$  parametri, joka vastaa alkuperäisen jakauman  $F$  parametria  $\theta$ . Tehdään havaitusta otoksesta  $n$ -alkion otos palauttaen. Tätä otosta kutsutaan bootstrap-otokseksi ja merkitään  $y_1^*, y_2^*, \dots, y_n^*$ . Lasketaan bootstrap-otoksesta estimaatti  $\hat{\theta}^*$  populaation parametrille  $\theta$  samalla tavalla kuin alkuperäisen aineiston tapauksessa lasketaan  $\hat{\theta}$ . Toistetaan tätä  $B$ -kertaa ja saadaan  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ , jota kutsutaan estimaattorin  $\hat{\theta}$  bootstrap-jakaumaksi, sillä se määräytyy empiirisen kertymäfunktion  $\hat{F}$  perusteella.

Bootstrap-jakauman odotusarvon estimaattori on

$$(6.1) \quad \bar{\theta}^* = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^*.$$

Lisäksi estimaattori bootstrap-jakauman varianssille on

$$(6.2) \quad \hat{\sigma}^2(\hat{\theta}^*) = \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^* - \bar{\theta}^*)^2.$$

Jos voidaan olettaa, että yhteys  $F$ :n ja  $\theta$ :n välillä on samanlainen kuin  $\hat{F}$ :n ja  $\hat{\theta}$ :n välillä siten, että erotuksen  $\hat{\theta} - \theta$  jakauma on likimain sama kuin boot-

strap-jakaumasta määritetyn erotuksen  $\hat{\theta}^* - \hat{\theta}$  jakauma, voidaan osoittaa, että seuraava tulos on voimassa

$$\text{Var}(\hat{\theta}) = \text{Var}_F[\hat{\theta} - \theta] \approx \text{Var}_{\hat{F}}[\hat{\theta}^* - \hat{\theta}] = \hat{\sigma}^2(\hat{\theta}^*).$$

Lisäksi estimoiduksi harhaksi saadaan

$$E_F[\hat{\theta} - \theta] \approx E_{\hat{F}}[\hat{\theta}^* - \hat{\theta}] = E_{\hat{F}}[\hat{\theta}^*] - \hat{\theta} \approx \bar{\theta}^* - \hat{\theta},$$

joten harhakorjattu estimaattori parametrille  $\theta$  on

$$\hat{\theta} - (\bar{\theta}^* - \hat{\theta}) = 2\hat{\theta} - \bar{\theta}^*.$$

(Garthwaite, Jolliffe & Jones, 2002.)

Bootstrap-jakaumasta  $\hat{\theta}^*$  lasketaan usein myös luottamusväli populaation parametrille  $\theta$ . Luottamusväli voidaan määrittää monella eri tavalla taustalla olevista oletuksista riippuen (Davison et al. 1997). Tässä esitellään vain yksi usein käytetty perus-bootstrap-väli, jonka taustalla on sama oletus kuin yllä. Määritellään  $100(1 - 2\alpha)$  % perus-bootstrap-väli  $\theta$ :lle.

$$\left( \hat{\theta} - (\theta^*(1 - \alpha) - \hat{\theta}), \hat{\theta} + (\hat{\theta} - \theta^*(\alpha)) \right).$$

missä  $\theta^*(\alpha)$  tarkoittaa tason  $\alpha$  kvantiilia bootstrap-jakaumalle eli  $P_{\hat{F}}(\hat{\theta}^* \leq \theta^*(\alpha)) = \alpha$ .

### 6.1.2 Simulointikokeiden perusidea

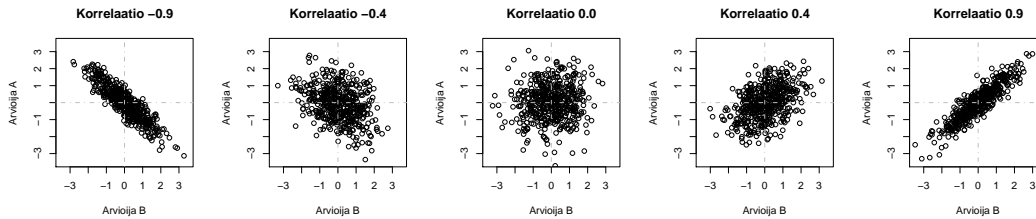
Simulointikokeissa aineisto muodostetaan siten, että generoidaan haluttu määrä havaintoja nollakeskisestä kaksiuotteisesta normaalijakaumasta, kun otoskoko, varianssi ja korrelaatiokerroin on annettu. Tarkemmin simuloidaan  $n$  havaintoa jakaumasta

$$\mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Kaksiuotteinen normaalijakauma soveltuu tilanteeseen hyvin, sillä korrelaatiokertoimen avulla voidaan arvioida yksimielisyyden voimakkuutta populaatiossa. Lisäksi kaksiuotteinen normaalijakauma on symmetrinen odotusarvon suhteen, mikä osoittautuu hyödylliseksi ominaisuudeksi, jotta arviot saadaan muunnettua kaksiluokkaisiksi.

Generoitaessa kaksiuotteisesta normaalijakaumasta haluttu määrä havaintoja saadaan kaksi muuttujaa, joista ensimmäinen viittaa arvioijan A ja toinen arvioijan B arvioihin. Tilastoyksikkönä on luonnollisesti arvioitava yksikkö kuten sairas henkilö, tuen hakija tai muu vastaava. Kaksiuotteisen normaalijakauman symmetrisyydestä nollan suhteen seuraa, että aineisto voidaan luokitella siten, että positiivinen generoitu havainto tulkitaan arvioijan myönteiseksi vastaukseksi ja negatiivinen arvo kielteiseksi vastaukseksi kyseisen yksikön kohdalla. Näin aineisto vastaa tilannetta, jossa on kaksi päätösvaihtoehtoa.

Simuloidusta otoksesta lasketaan kapan estimaatti  $\hat{\kappa}$  tai jokin muu estimoitava tunnusluku kuten hajonta tai Cohenin kapan ja modifioidun kapan erotus, minkä jälkeen otoksesta tehdään tuhat bootstrap-otosta palauttaen. Jokaisesta bootstrap-otoksesta lasketaan kapan estimaatti  $\hat{\kappa}_i^*$ . Näin saadaan bootstrap-jakauma  $\hat{\kappa}^*$ , josta lasketaan bootstrap-jakauman odotusarvon estimaatti. Tätä toistetaan erilaisilla korrelaatiokertoimilla ja otoksilla, jotta voidaan tarkastella niiden vaikutusta tarkasteltavaan tunnuslukuun.



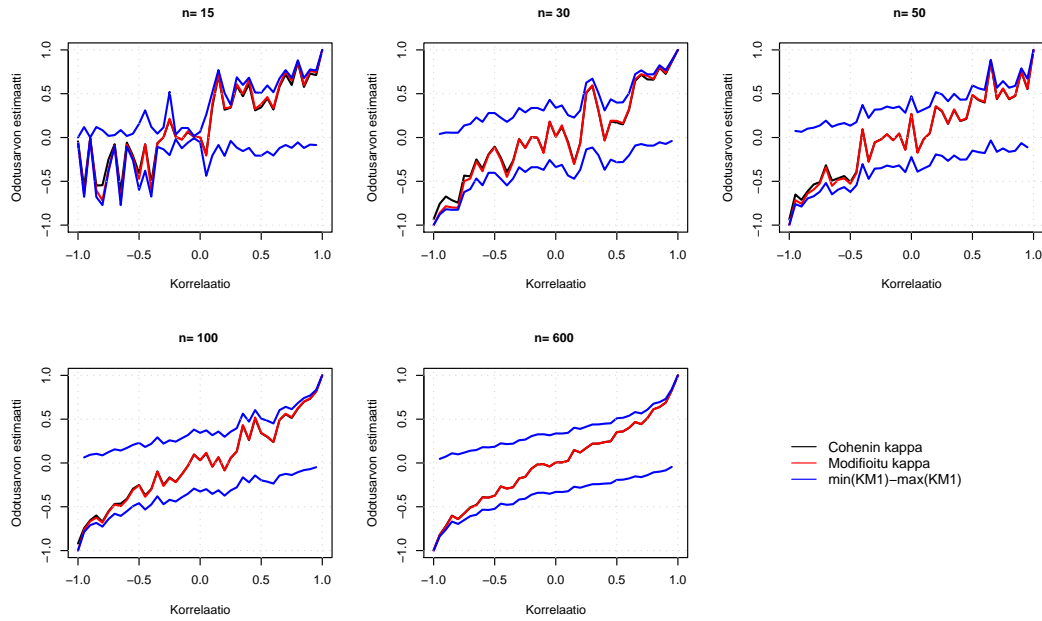
**Kuvio 6.1.** Kaksiulotteinen normaalijakauma eri korrelaatiokertoimilla.

Kuvio 6.1 havainnollistaa aineiston simulointia. Huomataan, että korrelaatiokertoimen kasvaessa ellipsi kapenee, joten aineistoon tulee enemmän havaintoja, joissa arvioijat ovat yksimielisiä. Päinvastaisesti korrelaatiokertoimen ollessa todella pieni, generoituu aineistoon suurimmaksi osaksi havaintoja, joissa arvioijat ovat erimielisiä. Tärkeä huomio on se, että kaikissa tilanteissa tulee lähes sama määrä ei-ei ja kyllä-kyllä -vastauksia sekä vastaavasti käy vastauksien ei-kyllä ja kyllä-ei kanssa. Aikaisempien tuloksien perusteella tiedetään, että kontingenssitaulukon rakenne vaikuttaa Cohenin kapan ja modifioidun kapan arvoon ja, että vastaavanlaisissa tilanteissa ne toimivat hyvin.

## 6.2 Kappojen vertailu

Tarkastellaan Cohenin kapan ja modifioidun kapan käyttäytymistä korrelaatiokertoimen funktiona. Hyödynnetään tarkastelussa bootstrap-jakaumien odotusarvon estimaattia. Kuitenkin simuloiduissa aineistoissa päätösvaihtoehtojen jakauma on melko symmetrinen, joten vinot jakaumat tulee huomioida tarkastelussa erikseen. Kuviossa 6.2 havainnollistetaan tuloksia graafisesti. Kuviossa musta viiva kuvaa Cohenin kapan bootstrap-jakauman odotusarvon estimaattia korrelaatiokertoimen funktiona ja vastaavasti punainen viiva modifioidun kapan vastaavia estimaatteja. Sinisellä viivalla sen sijaan kuvataan modifioidun kapan minimiä ja maksimiä, kun havaittu yksimielisyys  $\theta_0$  on vakioitu (lauseke 3.5). Tarkastelussa käytetään modifioidun kapan ääriarvoja, sillä modifioidun kapan maksimi on suurempi tai yhtä suuri kuin Cohenin kapan maksimi, mutta niiden minimi ovat yhtä suuria. Kaikissa kuvaajissa korrelaatiokertoimen tarkkuus on 0.05, mutta otoskoko vaihtelee eri kuvaajissa 15–600 havaintoon.

Kuviosta 6.2 havaitaan, että kaikilla otoksilla Cohenin kappaa ja modifioitua kappa ovat likimain samansuuruisia. Tämä selittyy päätösvaihtoehtojen jakau-



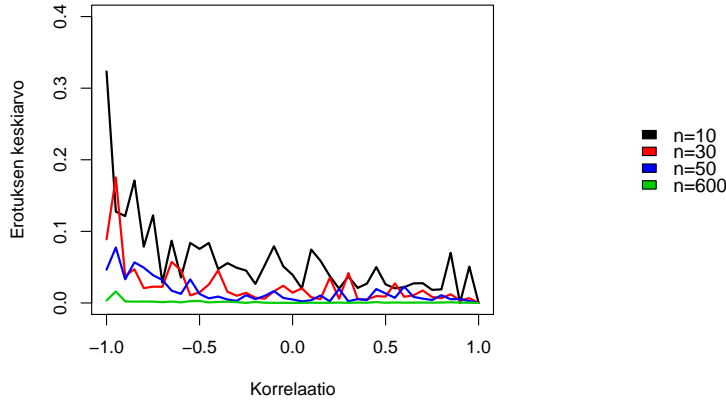
**Kuvio 6.2.** Cohenin kappan ja modifioidun kappan bootstrap-jakauman odotusarvon estimaatti korrelaatiokertoimen funktiona.

man symmetrisyydellä, sillä tästä syystä simuloiduissa aineistoissa ei-diagonaalien erotus  $\hat{\pi}_{12} - \hat{\pi}_{21}$  jää pieneksi, joten myös kappojen ero on pieni (lause 3.4). Huomataan myös ominaisuus, että otoskoon kasvaessa kappojen odotusarvojen estimaattien käyrä tasoittuu. Tätä ominaisuutta voidaan perustella siten, että pienillä otoksilla estimointiin liittyvä harha on suuri, mistä syystä odotusarvojen estimaatit vaihtelevat suuresti. Pienillä otoksilla on myös suurempi riski saada päätösvaihtoehdoille vino jakauma siinäkin tilanteessa, että perusjoukossa vastaava jakauma olisi melko symmetrinen. Ongelma voidaan välttää suunnittelemalla tutkimus siten, että otos on riittävän suuri (vertaa luku 5). Jos kuitenkin havaintoja on vähän (esim. alle 20), tulee tuloksia tulkita harkiten.

Modifioidun kappan bootstrap-jakauman odotusarvon estimaatin minimin ja maksimin erotus, kun havaittu yksimielisyys on kiinnitetty, pysyy korrelaatiokertoimesta riippumattomasti kutakuinkin vakiona. Suurimmillaan erotus on kuitenkin korrelaatiokertoimen ollessa lähellä  $\pm 1$ . Yllättävä havainto on se kuinka laaja väli minimin ja maksimin välillä on korrelaatiokertoimesta riippumatta. Tämä vahvistaa tuloksen, että joissain tilanteissa estimoidun kappan arvo aliarvioi todellisen yksimielisyyden tason. Tulee kuitenkin muistaa, että nimenomaisesti vinot päätösvaihtoehtojen jakaumat aiheuttavat poikkeavan pienet havainnot. Toisin sanoen tutkimusongelma tulisi mahdollisuuksien mukaan asettaa siten, että päätösvaihtoehtojen jakaumasta ei muodostuisi todella vinoa.

Tutkitaan vielä tarkemmin modifioidun kappan ja Cohenin kappan itseisarvon erotusta  $|\kappa_{M1}| - |\kappa_C|$ . Hyödynnetään jälleen Bootstrap-menetelmää, mutta estimoitava parametri on kappojen itseisarvojen erotus  $|\hat{\kappa}_{M1}| - |\hat{\kappa}_C|$ . Kuvios-

ta 6.3 voidaan huomata, että korrelaatiokertoimen ollessa negatiivinen erotus on suurimmillaan, kun taas voimakkaalla positiivisella korrelaatiolla erotus lähestyy nollaa. Yllättävä havainto on, että suurilla otoksilla erotuksesta muodos-



**Kuvio 6.3.** Kappojen itseisarvojen erotus  $|\kappa_{M1}| - |\kappa_C|$  korrelaatiokertoimen funktiona.

tuu todella pieni korrelaatiokertoimesta huolimatta. Tämä selittyy simuloidun aineiston ominaisuuksilla. Nimittäin suurilla otoksilla tarkasteluun tulevilla aineistoilla ei-diagonaalialkioiden erotus  $\pi_{12} - \pi_{21}$  jää todella pieneksi päätösvaihtoehtojen jakauman symmetrisyyden takia. Näin ollen lauseen 3.4 mukaan kappojen nimittäjien ero on pieni, joten kappojen erotuskin jää suhteessa pieneksi. Tulee kuitenkin pitää mielessä, että kyse on nimittäjien erotuksesta, eikä se kerro suoraan kappojen eron suuruutta. Käytännön sovelluksissa kappoja käytetään useimmiten melko symmetristen päätösvaihtoehtojen tapauksessa, sillä silloin niiden tulkinta on johdonmukainen. Toisin sanoen käytännön sovelluksissa kappojen erotus jää usein pieneksi.

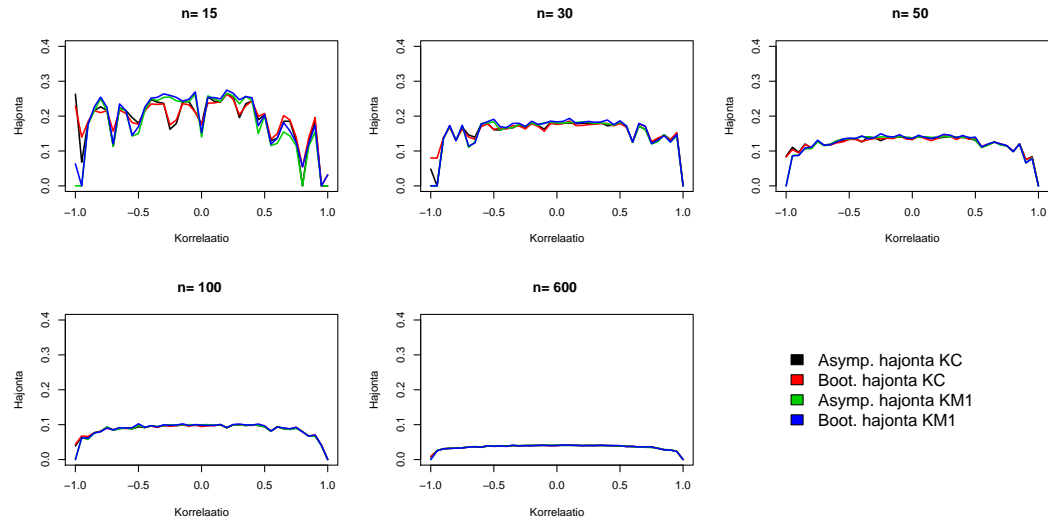
### 6.3 Hajontojen vertailu

Tutkitaan seuraavaksi hajonnan riippuvuutta korrelaatiokertoimesta  $\rho$ . Tässäkin tarkastelussa aineisto simuloidaan kaksiulotteisesta normaalijakaumasta, jonka jälkeen siitä määritellään Cohenin kapan ja modifioidun kapan asymptoottiset varianssit (lausekkeet 4.6 ja 4.8). Tämän jälkeen hyödynnetään Bootstrap-menettelmää, ja lasketaan hajonnan odotusarvon estimaatti.

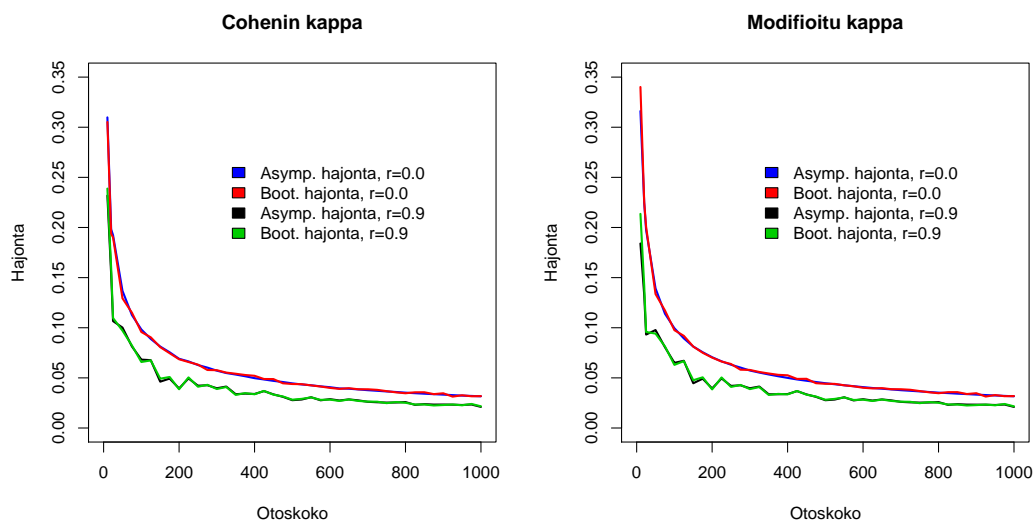
Kuviossa 6.4 havainnollistetaan tilannetta graafisesti. Huomataan, että suurilla otoksilla hajonta pysyy lähes samana korrelaatiokertoimesta riippumatta lukuun ottamatta tilanteita, joissa korrelaatiokerroin on lähellä  $\pm 1$ . Pienillä otoksilla hajontojen vaihtelu on todella suurta, mutta ne tasoittuvat otoskoon kasvaessa. Tärkeä havainto on se, että kaikki hajonnat ovat kutakuinkin samansuuruisia. Täten asymptoottiset hajonnat näyttäisivät toimivan hyvin ainakin



tarkasteluun mukaan tulevissa tilanteissa. Toisaalta muunnoksesta huolimatta modifioidulla kapalla hajonnan suuruusluokka säilyy kutakuinkin samana kuin Cohenin kapalla. Eri kuvioita verratessa havaitaan, että hajonta pienenee otoskoon funktiona, niin kuin oletettua onkin.



**Kuvio 6.4.** Cohenin kaplan ja modifioidun kaplan suurten otosten hajonnat ja hajontojen bootstrap-jakaumien odotusarvojen estimaatit korrelaatiokertoimen funktiona.



**Kuvio 6.5.** Cohenin kaplan ja modifioidun kaplan hajontojen vertailu otoskoon funktiona.

Tutkitaan hajontaa otoskoon funktiona siten, että x-akselilla on otoskoko ja erivärisillä käyrillä kuvataan korrelaatiokertoimen suuruutta (kuvio 6.5).

Korrelaatiokertoimiksi kuvaajaan valitaan  $\rho = 0$  ja  $\rho = 0.9$ , sillä riippuvuus on samanlainen negatiivisilla korrelaatioilla, ja avoimella välillä  $(0, 0.9)$  olevien korrelaatiokertoimien käyrät asettuvat kuvaajassa olevien käyrien väliin.

Kuten jo edellisen kuvion 6.4 tapauksessa havaittiin, korrelaatiokerroin ei vaikuta hajontaan kovinkaan paljon. Kuitenkin hajonta on pienempää suurilla korrelaatiokertoimilla. Kiinnostavampi ominaisuus on se, että hajonta laskee aluksi todella jyrkästi otoskoon funktiona, kunnes noin 150 havainnon kohdalla käyrän kulmakerroin loivenee merkittävästi. Toisin sanoen yhden havainnon lisäys tiputtaa hajonnan suuruutta huomattavasti vähemmän kuin pienemmällä otoksilla. Muun muassa tästä syystä otoskoon määrittäminen ennalta määrätyn luottamusvälin leveyden tai testauksen voimakkuuden avulla, ohjaa valitsemaan optimaalisen otoskoon. Näin ollen tutkimukseen kohdistettavat resurssit saadaan hyödynnettyä tehokkaasti.

## 7 Yhteenveto

Tutkielma aloitettiin Cohenin kappan ominaisuuksien ja tulkinnan esittelyllä. Kävi ilmi, että Cohenin kappaa toimii epäjohdonmukaisesti muun muassa tilanteissa, joissa esiintyvyys  $\theta_P$  on suuri tai Cohenin kappaa  $\kappa_C$  ollessa negatiivinen. Näissä tilanteissa Cohenin kappaa arvo vääristyy ja saattaa siten johtaa vääriin johtopäätelmiin. Tästä syystä monet tutkijat ovat esittäneet parannusehdotuksia Cohenin kappalle. Tässä tutkielmassa tarkasteltiin yksityiskohtaisesti muutamia näistä muunnoksista.

Aluksi esiteltiin modifioitu kappaa, jonka tavoitteena on parantaa Cohenin kappaa tulkittavuutta erimielisyyden tilanteessa. Havaittiin, että modifioitu kappaa säilyttää Cohenin kappaa hyvät ominaisuudet ja, että se muuttaa Cohenin kappaa oikeaan suuntaan. Osoitettiin myös tulos, että modifioitu kappaa on itseisarvoltaan aina suurempi tai yhtä suuri kuin Cohenin kappaa. Toisin sanoen modifioitu kappaa yliarvioi yksimielisyyden tason Cohenin kappaa nähden. Simulointikokeiden perusteella havaittiin kuitenkin, että käytännön sovelluksissa erotus jää usein pieneksi. Kuitenkaan modifioitu kappaa ei korjaa Cohenin kappaa epäjohdonmukaista käyttäytymistä tilanteissa, joissa päätösvaihtoehtojen jakauma on todella vino. Näin ollen esitettiin vielä standardointiin perustuvia muunnoksia, joita sovelletaan Cohenin kappaa ja modifioituun kappaa.

Standardointi perustuu siihen havaintoon, että jonkin tekijän ollessa vakio kappaa maksimi ja minimi muuttuvat vakioitavan tekijän funktiona. Tästä syystä standardoinnin ideana on yhteismitallistaa kappaa arvot välille  $[0, 1]$ , kun havaittu yksimielisyys  $\theta_0$  tai sattumayksimielisyys  $\theta_E$  on vakioitu. Todella yllättävä tulos oli, että standardoitujen kappaa tulkinta muuttuu vakioinnin seurauksena, eivätkä ne siten mittaa yksimielisyyttä. Toki niitä voidaan hyödyntää tulkintaa tehtäessä, mutta on kyseenalaista, ovatko ne tarpeellisia, sillä samat päätelmät voidaan tehdä muutenkin.

Koska yksikäsitteistä mittaa yksimielisyydelle ei löytynyt, nousee kysymys, miten käytännössä tulisi toimia. Ensinnäkin ennen tutkimusta kannattaa arvioida riittävä otoskoko Cohenin kappaa perustuvalla menetelmällä. Näin varmistetaan, että esimerkiksi väliestimoinnissa savutetaan riittävä tarkkuus. Toiseksi tulkinnassa tulee huomioida siihen vaikuttavat tekijät kuten arvioijien ammattitaito, kontingenssitaulukon rakenne, otoskoko ja hajonta. Lisäksi yksimielisyyden tason riittävyttä arvioitaessa tulee luonnollisesti huomioida tutkittavan asian vakavuus ja luokitteluperusteiden tarkkuus. Koska tulkintaan vaikuttavat monet tekijät, tulee myös raportointiin kiinnittää erityistä tarkkuutta.

Tulevaisuudessa olisi mielenkiintoista tutkia vielä tarkemmin Cohenin kappan ja modifioidun kappan käyttäytymistä erikoistapauksissa simulointikokeiden avulla. Varsinkaan modifioitua kappaa ei ole juuri tutkittu simulointikokeilla, joten ne voisivat antaa uusia mielenkiintoisia tuloksia. Lisäksi olisi hyödyllistä löytää yleinen tapa tulkita kappaa erikoistilanteissa ja pohtia tarkasti, mitkä tekijät tulee huomioida ja miten. Lisäksi tämän tutkimuksen tarkasteluja voisi laajentaa myös useammalle kuin kahdelle luokalle ja sen jälkeen järjestysasteikolliselle luokitukselle sekä useammalle arvioijalle. Uusia mielenkiintoisia tutkimusongelmia ja kehitysideoita on siis todella paljon.

# Lähdeluettelo

- Agresti, A. (2007), "An Introduction to Categorical Data Analysis", *Wiley-Interscience*.
- Banerjee, M., Capozzoli, M., McSweeney L. & Sinha, D. (1999), "Beyond Kappa: A Review of Interrater Agreement Measures", *The Canadian Journal of Statistics*, 27, 3–23.
- Bloch, D. A. & Kraemer, H. C. (1989), "2 x 2 Kappa Coefficients: Measures of Agreement or Association", *Biometrics*, 45, 269–287.
- Cantor, A. B. (1996), "Sample-Size Calculations for Cohen's Kappa", *Psychological Method*, 1, 150–153.
- Casella, G. & Berger, R. L. (2002), "Statistical Inference Second Edition", *Duxbury*.
- Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales", *Educational and Psychological Measurement*, 20, 37–46.
- Cohen, J. (1968), "Weighted kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit", *Psychological Bulletin*, 70, 213–220.
- Davison, A. C. & Hinkley, D. V. (1997), "Bootstrap Methods and their Application", *Cambridge University Press*.
- Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, 7, 1–26.
- Everitt, B. S. (1968), "Moments of the Statistics Kappa and Weighted Kappa", *The British Journal of Mathematical and Statistical Psychology*, 21, 97–103.
- Flack, V. F., Affi, A. A. & Lachenbruch, P. A. (1988), "Sample Size Determinations for the two Rater Kappa Statistic", *Psychometrika*, 53, 321–325.
- Fleiss, J. L. (1971), "Measuring Nominal Scale Agreement Among Many Raters", *Psychological Bulletin*, 76, 378–382
- Fleiss, J. L., Cohen, J. & Everitt, B. S. (1969), "Large Sample Standard Errors of Kappa and Weighted kappa", *Psychological Bulletin*, 72, 323–327.
- Garthwaite, P. H., Jolliffe, I. T. & Jones, B. (2002), "Statistical Inference", *Oxford University Press*.
- Goodman, L. A. & Kruskal, W. H. (1954), "Measures of Association for Cross Classifications", *Journal of the American Statistical Association*, 49, 732–764.
- Hoehler, F. K. (2000), "Bias and Prevalence Effects on Kappa Viewed in Terms of Sensitivity and Specificity", *Journal of Clinical Epidemiology*, 53, 499–503.
- Hogg, R. V. & Tanis E. A. (2006), "Probability and Statistical Inference", *Upper Saddle River (N.J.), Pearson Prentice Hall*
- Johnson, N. L. & Kotz S. (1969)"Discrete Distributions : Distributions in Statistics", *New York : Houghton Mifflin*.

- Korpela, K. M. & Ylén, M. (2007), "Perceived Health is Associated with Visiting Natural Favourite Places in the Vicinity", *Health & Place*, 13, 138–151.
- Landis, J. R. & Koch, G. G. (1977), "The Measurement of Observer Agreement for Categorical Data", *Biometrics*, 33, 159–174.
- Light, R. J. (1971), "Measures of Response Agreement for Qualitative Data: Some Generalizations and Alternatives", *Psychological Bulletin*, 76, 365–377.
- Pornpis, Y. (2007), "Some Statistical Aspects of Assessing Agreements: Theory and Applications", *The Degree of Doctor of Philosophy (Mathematics) Faculty of Graduate Studies Mahidol University, Bangkok, Thailand*.
- Salas, S. L., Hille, E. & Etgen, G. (2003), "Calculus - Ninth Edition", *John Wiley & Sons, Inc*, 663–680.
- Sim, J. & Wright, C. C. (2005), "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements", *Physical Therapy*, 85, 257–268.
- Sinha, B. K., Pornpis, Y. & Montip T. (2006), "Cohen's kappa statistic: a critical appraisal and some modifications", *Calcutta Statistical Association Bulletin*, 58, 151–169.

## Liite A: Maksimaalinen Q, kun Cohenin kappa tuntematon

$\pi_1$	$\pi_2$	Q	$\kappa_C$
0.1	0.1	2.21417	0.366
0.2	0.1	1.44128	0.339
0.2	0.2	1.31200	0.289
0.3	0.1	0.93136	0.310
0.3	0.2	1.07992	0.243
0.3	0.3	1.07003	0.187
0.4	0.1	0.58424	0.256
0.4	0.2	0.84096	0.177
0.4	0.3	0.97355	0.121
0.4	0.4	1.00558	0.067
0.5	0.1	0.36000	0.000
0.5	0.2	0.64000	0.000
0.5	0.3	0.84000	0.000
0.5	0.4	0.96000	0.000
0.5	0.5	1.00000	0.000
0.6	0.1	0.25684	0.000
0.6	0.2	0.48980	0.000
0.6	0.3	0.69136	0.000
0.6	0.4	0.85207	0.000
0.6	0.5	0.96000	0.000
0.6	0.5	0.96000	0.000
0.7	0.1	0.17355	0.000
0.7	0.1	0.17355	0.000

$\pi_1$	$\pi_2$	Q	$\kappa_C$
0.7	0.2	0.34964	0.000
0.7	0.3	0.52438	0.000
0.7	0.4	0.69136	0.000
0.7	0.5	0.84000	0.000
0.7	0.5	0.84000	0.000
0.8	0.2	0.22145	0.000
0.8	0.3	0.34964	0.000
0.8	0.4	0.48980	0.000
0.8	0.5	0.64000	0.000
0.8	0.5	0.64000	0.000
0.8	0.5	0.64000	0.000
0.8	0.5	0.64000	0.000
0.9	0.1	0.04819	0.000
0.9	0.2	0.10519	0.000
0.9	0.3	0.17355	0.000
0.9	0.4	0.25684	0.000
0.9	0.5	0.36000	0.000
0.9	0.5	0.36000	0.000
0.9	0.5	0.36000	0.000
0.9	0.5	0.36000	0.000

## Liite B: Parametrin Q maksimi, kun Cohenin kapa arvioitu

		Cohenin kappa									
$\pi_1$	$\pi_1$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.1	1.000	1.598	1.984	2.179	2.205	2.083	1.835	1.481	1.043	0.542
0.4	0.1	0.490	0.550	0.580							
0.4	0.2	0.793	0.832	0.840	0.819	0.771	0.697				
0.4	0.3	0.953	0.973	0.965	0.929	0.867	0.780	0.668	0.533		
0.4	0.4	1.000	1.004	0.984	0.940	0.872	0.781	0.668	0.533	0.376	0.198
0.5	0.1	0.360	0.356								
0.5	0.2	0.640	0.634	0.614	0.582						
0.5	0.3	0.840	0.832	0.806	0.764	0.706	0.630	0.538			
0.5	0.4	0.960	0.950	0.922	0.874	0.806	0.720	0.614	0.490	0.346	
0.5	0.5	1.000	0.990	0.960	0.910	0.840	0.750	0.640	0.510	0.360	0.190
0.6	0.1	0.257	0.207								
0.6	0.2	0.490	0.448	0.408							
0.6	0.3	0.691	0.659	0.621	0.576	0.524					
0.6	0.4	0.852	0.830	0.796	0.748	0.686	0.610	0.519			
0.6	0.5	0.960	0.950	0.922	0.874	0.806	0.720	0.614	0.490	0.346	
0.7	0.1	0.174									
0.7	0.2	0.350	0.280								
0.7	0.3	0.524	0.472	0.424	0.379						
0.7	0.4	0.691	0.659	0.621	0.576	0.524					
0.7	0.5	0.840	0.832	0.806	0.764	0.706	0.630	0.538			
0.8	0.1	0.105									
0.8	0.2	0.221	0.129								
0.8	0.3	0.350	0.280								
0.8	0.4	0.490	0.448	0.408							
0.8	0.5	0.640	0.634	0.614	0.582						
0.9	0.1	0.048									
0.9	0.2	0.105									
0.9	0.3	0.174									
0.9	0.4	0.257	0.207								
0.9	0.5	0.360	0.356								