Luku 13

Hypoteesien testaus

Hypoteesin H_0 testaus on menettely, jolla arvioidaan havaintojen tarjoamaa todistusvoimaa hypoteesia vastaan tai sen puolesta. Olkoon H_0 jonkin asiantilaa koskeva oletus eli hypoteesi. Teemme sitten havaintoja ja tarkistamme, onko tämä oletus sopusoinnussa havaintojen kanssa. Menettely muistuttaa jossain määrin matematiikasta tuttua todistustekniikkaa vastaoletuksen avulla. Tilastollinen hypoteesi H_0 ei tavallisesti johda havaintojen ja hypoteesin väliseen loogiseen ristiriitaan, mutta havainnot saattavat olla äärimmäisen epätodennäköisiä, mikäli H_0 olisi tosi. Mitä epätodennäköisempiä saadut havainnot ovat H_0:n vallitessa, sitä vahvemmin niiden katsotaan osoittavan, että H_0 ei pidä paikkansa.

Esimerkki 13.1 Heitetään lanttia 100 kertaa. Kruunujen lukumäärä X noudattaa binomijakaumaa $\mathrm{Bin}(100, \theta)$. Testataan, onko raha harhaton. Testattava hypoteesi on siis H_0: $\theta = \frac{1}{2}$.

Jos H_0 pitää paikkansa, niin

$$f(x) = \binom{100}{x}\left(\frac{1}{2}\right)^{100}, \qquad x = 0, 1, \ldots, 100.$$

Tiedämme, että $E(X) = 50$ ja odotusarvon läheisyydessä olevat X:n arvot ovat kaikkein todennäköisimpiä. Voimme valita *testisuureeksi* esimerkiksi satunnaismuuttujan $T = |X - 50|$, joka mittaa, kuinka hyvin havainnot ja hypoteesi sopivat yhteen.

Jos havaitaan esimerkiksi $X = 30$, niin $T = |30 - 50| = 20$. Jos H_0 on tosi, on todennäköisyys saada näin poikkeava tai vielä poikkeavampi arvo

$$P(T \geq 20) = P(|X - 50| \geq 20) \approx 0.$$

On siis äärimmäisen epätodennäköistä, että harhattomalla lantilla saadaan näin poikkeava tulos. Havainto tukee voimakkaasti käsitystä, että H_0 on väärä ja lantti on harhainen. $\square$

13.1 Testisuureet ja p-arvot

Merkitsevyyden testauksessa mahdolliset koetulokset (havainnot) on kyettävä asettamaan järjestykseen sen mukaan, miten hyvin ne sopivat yhteen hypoteesin kanssa. Tätä tarkoitusta varten valitaan jokin *testisuure* T, joka mittaa havaintojen ja hypoteesin H_0 yhteensopivuutta. Testisuureen T pieni arvo osoittaa havaintojen hyvää yhteensopivuutta hypoteesin kanssa ja T:n suuri arvo osoittaa huonoa yhteensopivuutta. Testisuure valitaan ennen kuin tutkitaan havaintoja. Valintaan vaikuttaa tietysti se, millaisia poikkeamia halutaan tunnistaa.

Kun havainnot on saatu, voidaan laskea testisuureen arvo $T = t$. Sitten voidaan laskea todennäköisyys, että saadaan ainakin näin poikkeava T:n arvo, jos H_0 on tosi. Tätä todennäköisyyttä

$$\alpha_h = P(T \geq t \mid H_0 \text{ tosi}) \tag{13.1.1}$$

kutsutaan *p-arvoksi* tai *havaituksi merkitsevyystasoksi*. Jos α_h on hyvin pieni, niin se osoittaa, että sellaista havaintoa ei saataisi juuri koskaan, jos H_0 on tosi. Silloin havainnot tukevat vahvasti oletusta, että H_0 ei pidä paikkaansa.

Esimerkki 13.2 Henkilölle jaetaan satunnaisessa järjestyksessä pöytään kuvapuoli alaspäin 4 korttia, jotka ovat eri maata ($\spadesuit$, $\heartsuit$, $\clubsuit$, $\diamondsuit$). Jokaisen kortin kohdalla henkilö arvaa, mitä maata kortti on. Henkilö väittää, että hänellä on sellaisia yliaistillisia kykyjä, että hän pystyy keskittymällä "näkemään" kortin ja pääsemään parempiin tuloksiin kuin pelkkä arvaaja. Ajatellaan, että henkilöllä on 4 kirjekuorta, jotka on merkitty symboleilla $\spadesuit$, $\heartsuit$, $\clubsuit$ ja $\diamondsuit$. Kortit on pistettävä oikeisiin kirjekuoriin.

Olkoon X oikein sijoitettujen korttien lukumäärä. Jos henkilö on pelkkä arvaaja, niin silloin oikeiden lukumäärän X todennäköisyysfunktio on

X	0	1	2	4	Yhteensä
$f(x)$	$\frac{9}{24}$	$\frac{8}{24}$	$\frac{6}{24}$	$\frac{1}{24}$	1

Laskemalla voidaan todeta, että $E(X) = 1$ ja $\mathrm{Var}(X) = 1$. Testataan nyt nollahypoteesi H_0, että henkilö on arvaaja. Toistetaan koe 50 kertaa ja olkoon X_i oikein sijoitettujen korttien lukumäärä i. kokeessa, $i = 1, \ldots, 50$. Silloin jokainen satunnaismuuttuja X_i noudattaa edellä määriteltyä X:n jakaumaa. Satunnaismuuttujien summa

$$S = X_1 + X_2 + \cdots + X_{50}$$

noudattaa keskeisen rajaväittämän mukaan likimain normaalijakaumaa. Nyt $E(S) = 50 \cdot 1$ ja $\mathrm{Var}(S) = 50 \cdot 1$, joten

$$\frac{S - 50}{\sqrt{50}} \approx \mathrm{N}(0, 1).$$

Henkilö sai seuraavan tuloksen:

Oikein sijoitettujen lkm	0	1	2	4	Yhteensä
Havaittu frekvenssi	17	18	9	6	50

Aineistosta laskettu S:n arvo on 60. Siksi

$$\alpha_h = P(S \geq 60) \approx P\Big(Z \geq \frac{60-50}{\sqrt{50}}\Big) = 0.079,$$

missä Z noudattaa normaalijakaumaa $N(0,1)$. Koska p-arvo ei ole kovin pieni, niin H_0 saa jäädä voimaan. Toisaalta p-arvo on niin pieni, että joku saattaa epäillä H_0:n pätevyyttä. Eräs tapa yrittää selvittää epäilys, on kerätä lisää aineistoa, mikäli mahdollista. □

13.2 Testien arviointi

Tässä alaluvussa esitellään käsitteitä ja menetelmiä, joiden avulla voidaan vertailla keskenään eri testien hyvyyttä. Keskeisin käsite näissä tarkasteluissa on *testin voimakkuus*. Se luonnehtii testisuureen herkkyyttä tunnistaa vaihtoehtoinen hypoteesi. Hypoteesille H_0 (nollahypoteesi) asetetaan eksplisiittisesti *vaihtoehtoinen hypoteesi* H_1, joka valitaan sen mukaan, millaisia poikkeamia H_0:sta halutaan tunnistaa. *Testimenettely* (testaus) on sääntö, joka havaintojen perusteella valitsee toisen ja vain toisen hypoteeseista H_0 ja H_1.

Teknisesti hypoteesit konstruoidaan siten, että tehdään *parametriavaruuden* Θ *jako* kahteen osajoukkoon Θ_0 ja Θ_1 siten, että $\Theta = \Theta_0 \cup \Theta_1$ ($\Theta_0 \cap \Theta_1 = \emptyset$). Nyt siis $\Theta_1 = \Theta \setminus \Theta_0 = \Theta_0^c$, joten Θ_1 on Θ_0:n komplementti Θ:ssa. Näitä vaihtoehtoja merkitään tavallisesti

$$H_0 \colon \theta \in \Theta_0 \qquad \text{vastaan} \qquad H_1 \colon \theta \in \Theta_1.$$

Jos $\Theta = \mathbb{R}$ on reaalilukujen joukko, niin tyypillisiä hypoteeseja ovat esimerkiksi

$$\begin{cases} H_0 \colon \theta = 0 \\ H_1 \colon \theta \neq 0 \end{cases} \qquad \begin{cases} H_0 \colon \theta \leq 1 \\ H_1 \colon \theta > 1. \end{cases}$$

Ensimmäisessä tapauksessa $\Theta_0 = \{0\}$ on yksi piste (H_0 yksinkertainen) ja $\Theta_1 = \{\,\theta \mid \theta \neq 0\,\}$. Toisessa tapauksessa molemmat hypoteesit ovat ns. yhdistettyjä hypoteeseja eli ne sisältävät useita parametrin arvoja: $\Theta_0 = \{\,\theta \mid \theta \leq 1\,\}$, $\Theta_1 = \{\,\theta \mid \theta > 1\,\}$.

Testimenettelyssä otosavaruus S jaetaan *hylkäysalueeseen* ja *hyväksymisalueeseen*. Olkoon T valittu testisuure ja $x \in S$ jokin havainto. Testisuureen T arvon $T(x)$ perusteella valitaan joko H_0 tai H_1. Hylkäysalue $C \subset S$ määritellään siten, että H_0 hylätään kaikilla $x \in C$. H_0 hyväksytään, jos $x \notin C$. Hylkäysaluetta kutsutaan myös testin *kriittiseksi alueeksi*. Vastaavasti H_0:n hyväksymisalue on C:n komplementti C^c, joten $S = C \cup C^c$.

Testi ei ole kuitenkaan erehtymätön. Voi sattua, että $x \in C$, vaikka $\theta \in \Theta_0$. Silloin syntyy *I lajin virhe*: H_0 hylätään, vaikka se on tosi. I:n lajin virheen todennäköisyttä merkitään

$$P(H_0 \text{ hylätään}|\ H_0 \text{ tosi}) = \alpha.$$

Vastaavasti voi sattua, että $x \in C^c$, vaikka $\theta \in \Theta_1$. Tässä tapauksessa tehdään *II lajin virhe*: H_0 hyväksytään, vaikka se ei ole tosi. II:n lajin virheen todennäköisyyttä merkitään

$$P(H_0 \text{ hyväksytään}|\ H_0 \text{ ei tosi}) = \beta.$$

Tavoitteena on valita kriittinen alue siten, että virhetodennäköisyydet ovat mahdollisimman pienet

13.2.1 Testin voimakkuus

Testin keskeiset ominaisuudet voidaan luonnehtia *voimakkuusfunktion* avulla. Voimakkuusfunktio on

$$\begin{aligned}\gamma(\theta) &= P(X \in C;\, \theta)\\ &= P(H_0 \text{ hylätään, kun parametrin arvo on } \theta)\\ &= 1 - \beta(\theta),\end{aligned}$$

missä $\beta(\theta)$ on II:n lajin virheen todennäköisyys. Testin *merkitsevyystaso* α eli testin *koko* on

$$\alpha = \max_{\theta \in \Theta_0} \gamma(\theta),$$

joka on siis I lajin virheen maksimi H_0:n vallitessa. Kun H_0 on yksinkertainen eli $\Theta_0 = \{\theta_0\}$, niin $\alpha = P(X \in C;\, \theta_0)$.

Esimerkki 13.3 Olkoon X havainto normaalijakaumasta $\mathrm{N}(\theta, 1)$. Testataan havainnon $X = x$ avulla hypoteesit

$$\begin{aligned}H_0&\colon \theta \le 0,\\ H_1&\colon \theta > 0.\end{aligned}$$

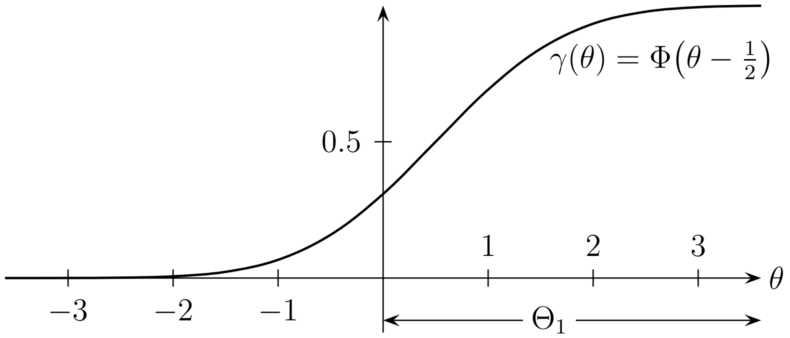
Olkoon kriittinen alue $C = \left\{\, x \;\middle|\; x \ge \tfrac{1}{2} \,\right\}$. Silloin testin voimakkuusfunktio on

$$\begin{aligned}\gamma(\theta) &= P\big(X \ge \tfrac{1}{2};\, \theta\big)\\ &= P\big(X - \theta \ge \tfrac{1}{2} - \theta;\, \theta\big)\\ &= 1 - \Phi\big(\tfrac{1}{2} - \theta\big) = \Phi\big(\theta - \tfrac{1}{2}\big),\end{aligned}$$

missä $X - \theta \sim \mathrm{N}(0, 1)$. Tässä tapauksessa γ on θ:n aidosti kasvava funktio ja $\gamma(\theta) \to 1$, kun $\theta \to \infty$ ja ja $\gamma(\theta) \to 0$, kun $\theta \to -\infty$. Testin merkitsevyystaso

$$\begin{aligned}\alpha &= \max_{\theta \in \Theta_0} \Phi\big(\theta - \tfrac{1}{2}\big)\\ &= \Phi\big(-\tfrac{1}{2}\big) \approx 0.31.\end{aligned}$$

□



Kuvio 13.1. Voimakkuusfunktio $\gamma(\theta)$, kun $X \sim \mathrm{N}(\theta, 1)$ ja kriittinen alue $C = \left\{\, x \mid x \geq \frac{1}{2} \,\right\}$.

Asettamamme tavoitteen mukaisesti pitäisi voimakkuuden $\gamma(\theta)$ olla mahdollisimman suuri, kun $\theta \in \Theta_1$. Vastaavasti voimakkuuden $\gamma(\theta)$ tulisi olla mahdollisimman pieni, kun $\theta \in \Theta_0$. Nämä vaatimukset ovat kuitenkin keskenään ristiriitaiset. Kun kriittistä aluetta kasvatetaan, kasvaa I lajin virhe. Kun hyväksymisaluetta kasvatetaan, kasvaa II lajin virhe. Eräs keino ratkaista tämä pulma, on kiinnittää ensin merkitsevyystaso α ja valita sitten näistä saman merkitsevyystason testeistä se, joka on voimakkain alueella $\theta \in \Theta_1$.

Tämän klassisen lähestymistavan esittivät Jerzy Neyman ja Egon S. Pearson. Neymanin ja Pearsonin lähestymistapakaan ei välttämättä johda yksikäsitteiseen ratkaisuun. Olkoot kahdella testillä voimakkuusfunktiot $\gamma_1(\theta)$ ja $\gamma_2(\theta)$ ja sama merkitsevyystaso. Silloin ei välttämättä pidä paikkansa, että $\gamma_1(\theta) \geq \gamma_2(\theta)$ kaikilla $\theta \in \Theta_1$ tai $\gamma_2(\theta) \geq \gamma_1(\theta)$ kaikilla $\theta \in \Theta_1$. Kumpikaan testi ei ole välttämättä tasaisesti voimakkaampi kuin toinen. Kuitenkin useissa tärkeissä testitilanteissa voidaan löytää tällainen *tasaisesti voimakkain testi.*

13.2.2 Testin konstruointi– yksinkertaiset hypoteesit

Olkoon X havainto jakaumasta, jonka tiheysfunktio on $f(x; \theta)$. Testataan nyt vaihtoehtoisia yksinkertaisia hypoteeseja

$$H_0\colon \theta = \theta_0 \qquad \text{ja} \qquad H_1\colon \theta = \theta_1. \tag{13.2.1}$$

Kun havainto $X = x$ on saatu, voidaan laskea *uskottavuussuhde*

$$\lambda(x) = \frac{L(\theta_0; x)}{L(\theta_1; x)},$$

jonka valitsemme testisuureeksi. Kun $\lambda(x)$ on suuri, olemme taipuvaisia hyväksymään H_0:n. Jos taas $\lambda(x)$ on pieni, valitsemme H_1:n.

Neymanin ja Pearsonin lähestymistavan mukaan kiinnitetään ensin merkitsevyystaso α. Tämän perusteella voimme valita λ:n kriittisen arvon siten, että

$$P[\lambda(X) \leq \lambda_\alpha;\ \theta_0] = \alpha. \tag{13.2.2}$$

Testin hylkäysalue on siis

$$C = \{\, x \mid \lambda(x) \le \lambda_\alpha \,\}. \tag{13.2.3}$$

Lause 13.1 (Neymanin ja Pearsonin apulause) *Kun testataan parametria θ koskevia hypoteeseja* (13.2.1)*, niin testi* (13.2.3) *on voimakkain kaikista merkitsevyystasoa α olevista testeistä. Merkitsevyystaso α on määritelty identiteetillä* (13.2.2)*.*

Todistus. Olkoon D mikä tahansa toinen kriittinen alue (toinen testi), jonka merkitsevyystaso on korkeintaan α. Silloin

$$\alpha = \int_C f(x;\theta_0)\,\mathrm{d}x \ge \int_D f(x;\theta_0)\,\mathrm{d}x,$$

joten

$$\int_{C\cap D^c} f(x;\theta_0)\,\mathrm{d}x \ \ge \int_{D\cap C^c} f(x;\theta_0)\,\mathrm{d}x. \tag{13.2.4}$$

Epäyhtälö (13.2.4) seuraa siitä, että

$$\int_{C\cap D^c} = \int_C - \int_{C\cap D} \qquad \text{ja} \qquad \int_{D\cap C^c} = \int_D - \int_{C\cap D}.$$

Jos $x \in C \cap D^c$, niin $x \in C$ ja silloin

$$f(x;\theta_1)\lambda_\alpha \ge f(x;\theta_0).$$

Jos taas $x \in D \cap C^c$, niin $x \in C^c$ ja

$$f(x;\theta_0) > f(x;\theta_1)\lambda_\alpha.$$

Siksi

$$\begin{aligned} \lambda_\alpha \int_{C\cap D^c} f(x;\theta_1)\,\mathrm{d}x \ &\ge \int_{C\cap D^c} f(x;\theta_0)\,\mathrm{d}x \\ &\ge \int_{D\cap C^c} f(x;\theta_0)\,\mathrm{d}x \ \ge\ \lambda_\alpha \int_{D\cap C^c} f(x;\theta_1)\,\mathrm{d}x, \end{aligned} \tag{13.2.5}$$

missä toiseksi viimeinen epäyhtälö on aito, elleivät C ja D ole samat. Kun epäyhtälö (13.2.5) jaetaan puolittain λ_α:lla ja epäyhtälön molemmille puolille lisätään integraali $\int_{C\cap D} f(x;\theta_1)\,\mathrm{d}x$, saadaan

$$\int_C f(x;\theta_1)\,\mathrm{d}x \ge \int_D f(x;\theta_1)\,\mathrm{d}x.$$

Koska määritelmän mukaan $\gamma_C(\theta_1) = \int_C f(x;\theta_1)\,\mathrm{d}x$ ja $\gamma_D(\theta_1) = \int_D f(x;\theta_1)\,\mathrm{d}x$, niin testi (13.2.3) on voimakkain kaikista testeistä, joiden merkitsevyystaso on korkeintaan α. $\square$

Esimerkki 13.4 Olkoon $X_1, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\mu, 1)$. Testataan kahta yksinkertaista hypoteesia H_0: $\mu = 1$ vastaan H_1: $\mu = 2$. Etsimme nyt parhaan kriittisen alueen C, kun testin merkitsevyystaso α on kiinnitetty. Määritämme ne otosavaruuden S pisteet, joissa

$$\lambda(x_1, \ldots, x_n) = \frac{L(1; x_1, \ldots, x_n)}{L(2; x_1, \ldots, x_n)} \leq \lambda_\alpha$$

ja $P[\lambda(X_1, \ldots, X_n) \leq \lambda_\alpha] = \alpha$. Koska $L(1) = c \cdot \exp\bigl[-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2\bigr]$ ja $L(2) = c \cdot \exp\bigl[-\frac{1}{2}\sum_{i=1}^n (x_i - 2)^2\bigr]$, niin

$$\log \lambda(x_1, \ldots, x_n) = -\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2 + \frac{1}{2}\sum_{i=1}^n (x_i - 2)^2 \leq \log \lambda_\alpha$$

Tästä seuraa, että

$$\sum_{i=1}^n (x_i - 2)^2 - \sum_{i=1}^n (x_i - 1)^2 \leq 2 \log \lambda_\alpha. \tag{13.2.6}$$

Koska $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \overline{x})^2 + n(\overline{x} - \mu)$, niin

$$\sum_{i=1}^n (x_i - 2)^2 - \sum_{i=1}^n (x_i - 1)^2 = n(\overline{x} - 2)^2 - n(\overline{x} - 1)^2 = n(-2\overline{x} + 3).$$

Siksi epäyhtälö (13.2.6) voidaan kirjoittaa muodossa

$$\overline{x} \geq -\frac{\log \lambda_\alpha}{n} + \frac{3}{2} = c_\alpha.$$

Koska H_0:n vallitessa $\overline{X} \sim \mathrm{N}(1, 1/\sqrt{n})$, niin määritetään c_α siten, että

$$P(\overline{X} \geq c_\alpha) = P[Z \geq \sqrt{n}(c_\alpha - 1)] = \alpha,$$

missä $Z = \sqrt{n}(\overline{X} - 1) \sim \mathrm{N}(0, 1)$ H_0:n vallitessa.

Oletetaan esimerkiksi, että $n = 16$ ja testin merkitsevyystaso $\alpha = 0.05$. Määritetään siis $c_{0.05}$ siten, että

$$P[Z \geq 4(c_{0.05} - 1)] = 0.05.$$

Silloin $4(c_{0.05} - 1) = 1.645$, joten

$$c_{0.05} = \frac{1.645}{4} + 1 = 1.41125.$$

Testin hylkäysalue merkitsevyystasolla $\alpha = 0.05$ on

$$C = \left\{ (x_1, \ldots, x_n) \;\middle|\; \overline{x} = \frac{1}{16}\sum_{i=1}^{16} x_i \geq 1.41125 \right\}.$$

□

13.3 Uskottavuussuhdetestit: Yksinkertaiset hypoteesit

Tilastollinen malli voidaan luonnehtia jakaumien joukkona. Jos jakaumalla on tiheysfunktio (tai todennäköisyysfunktio) niin parametrinen malli on tiheysfunktioden joukko

$$\mathcal{F} = \{ f(\,\cdot\,; \boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k \}, \tag{13.3.1}$$

missä jokainen annettu $\boldsymbol{\theta} \in \Theta$ määrittelee jonkin jakauman tiheysfunktion. Kun satunnaismuuttuja on diskreetti, tiheysfunktion paikalla on todennäköisyysfunktio. Seuraavassa ajatellaan, että testattava *tilastollinen hypoteesi* voidaan lausua tilastollisen mallin tuntemattomia parametreja koskevina oletuksina.

Hypoteesi H on *yksinkertainen*, jos se määrittää kaikkien tuntemattomien parametrien arvot niin, ettei malliin jää tuntemattomia parametreja. *Yhdistetty hypoteesi* ei sen sijaan täysin määritä tuntemattomien parametrien arvoja, mutta rajoittaa mahdollisten arvojen joukkoa.

13.3.1 Yksi parametri

Oletetaan aluksi, että mallissa on yksi tuntematon parametri θ. Testataan hypoteesia

$$H_0\colon \theta = \theta_0, \tag{13.3.2}$$

missä θ_0 on jokin θ:n numeerinen arvo.

Olkoon $l(\theta) = \log L(\theta)$ logaritmoitu uskottavuusfunktio ja $\hat{\theta}$ parametrin θ suurimman uskottavuuden estimaatti. Testataan hypoteesi (13.3.2) käyttäen uskottavuussuhteeseen perustuvaa Wilksin uskottavuustestisuuretta

$$\begin{aligned} W &= -2\log\frac{L(\theta_0)}{L(\hat{\theta})} \\ &= 2[l(\hat{\theta} - l(\theta_0))]. \end{aligned}$$

Testisuure $W = W(X_1, X_2, \ldots, X_n)$ on otoksen $X_1, X_2, \ldots, X_n$ funktio ja

$$W \xrightarrow{d} \chi^2(1),$$

kun n kasvaa. Jos testin riskitasoksi on valittu α, niin päätössääntö on:

$$\text{Hylkää } H_0 \text{ ja hyväksy } H_1, \text{ jos } W \geq \chi^2_\alpha(1),$$

missä $\chi^2_\alpha(1)$ on Khi2(1)-jakauman sellainen fraktiili, että $P[\chi^2 \geq \chi^2_\alpha(1)] = \alpha$.

Havainnoista x_1, x_2,, …, x_n voidaan laskea testisuureen arvo $W(x_1, x_2, \ldots, x_n) = d$. Kun arvo d on saatu, voidaan laskea testisuureen arvoon liittyvä todennäköisyys

$$\alpha_h = P(W \geq d \mid H \text{ tosi}),$$

joka on siis *havaittu merkitsevyystaso* eli *p-arvo*. Suurilla n:n arvoilla

$$P(W \geq d \mid H \text{ tosi}) \approx P(\chi^2(1) \geq d)$$

ja p-arvo voidaan siis määrittää Khi2(1)-jakauman avulla.

Esimerkki 13.5 Olkoon $X_1, X_2, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\theta, \sigma^2)$, missä varianssi σ^2 oletetaan tunnetuksi. Testataan hypoteesia

$$H_0\colon \theta = \theta_0 \text{ vaihtoehtoa } H_1\colon \theta \neq \theta_0 \text{ vastaan.}$$

Jos H_0 pitää paikkansa, uskottavuusfunktion arvo $L(\theta)$ on

$$L(\theta_0) = (2\pi\sigma^2)^{-n/2} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_0)^2]$$

ja

$$L(\hat{\theta}) = (2\pi\sigma^2)^{-n/2} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2].$$

Kun muodostetaan Wilksin testisuure $W = -2\log[L(\theta_0)/\, L(\hat{\theta})]$, saadaan

$$W = \frac{1}{\sigma^2} \sum_{i=1}^{n} [(x_i - \theta_0)^2 - (x_i - \bar{x})^2] = [\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma}]^2.$$

Silloin annetulla $d > 0$

$$W \geq d^2 \iff [\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma}]^2 \geq d^2,$$

joka toteutuu täsmälleen silloin kun

$$\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \leq -d \text{ tai } \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \geq d.$$

Koska H_0:n vallitessa $\frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \sim \mathrm{N}(0, 1)$, niin asettamalla ehto

$$P_{\theta_0}\,[\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \leq -d \text{ tai } \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \geq d] = \alpha$$

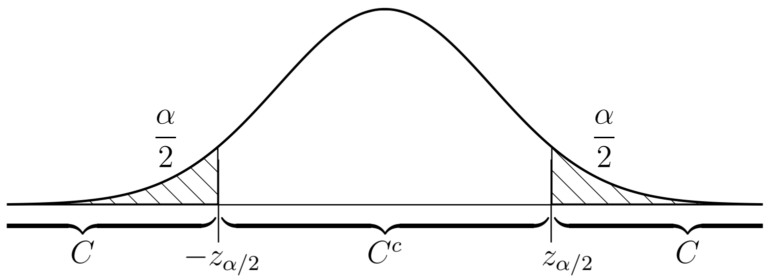
saadaan $d = z_{\alpha/2}$. Nyt uskottavuussuhdetesti on muotoa

$$\text{Testi:} \quad \begin{cases} H_0 \text{ hylätään, kun } |\frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma}| \geq z_{\alpha/2}; \\ \text{muutoin } H_0 \text{ jää voimaan.} \end{cases}$$

Nollahypoteesi H_0 hylätään, jos

$$|Z| = \left| \frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2},$$

joten testin kriittinen alue on siis $C = \{\, Z \mid |Z| \geq z_{\alpha/2} \,\}$.



Testin merkitsevyystaso on

$$\alpha = P(H_0 \text{ hylätään; } \theta = \theta_0) = P(|Z| \geq z_{\alpha/2};\, \theta = \theta_0).$$

Seuraavassa merkitään $P(Z \in C;\, \theta) = P_\theta(Z \in C)$. Voimakkuusfunktio on

$$\begin{aligned}
\gamma(\theta) &= P_\theta(Z \in C) \\
&= P_\theta(|Z| \geq z_{\alpha/2}) = P_\theta(Z \leq -z_{\alpha/2}) + P_\theta(Z \geq z_{\alpha/2}) \\
&= P_\theta\bigg(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \leq -z_{\alpha/2}\bigg) + P_\theta\bigg(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \geq z_{\alpha/2}\bigg) \\
&= P_\theta\bigg(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} + \frac{\theta - \theta_0}{\sigma/\sqrt{n}} \leq -z_{\alpha/2}\bigg) + P_\theta\bigg(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} + \frac{\theta - \theta_0}{\sigma/\sqrt{n}} \geq z_{\alpha/2}\bigg) \\
&= P_\theta\bigg(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \leq -z_{\alpha/2} - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\bigg) + P_\theta\bigg(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \geq z_{\alpha/2} - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\bigg) \\
&= \Phi\bigg(-z_{\alpha/2} - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\bigg) + 1 - \Phi\bigg(z_{\alpha/2} - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\bigg).
\end{aligned}$$

Merkitsevyystaso on

$$\gamma(\theta_0) = \Phi(-z_{\alpha/2}) + 1 - \Phi(z_{\alpha/2}) = 2\,\Phi(-z_{\alpha/2}) = 2 \cdot \frac{\alpha}{2} = \alpha.$$

Voimme todeta, että $\gamma(\theta) \to 1$, kun $\theta \to \infty$ tai $\theta \to -\infty$. □

Todennäköisyyden α_h tarkan arvon laskeminen usein hankala tehtävä, mutta sopivien ehtojen vallitessa likiarvo

$$\alpha_h = P(W \geq d \mid \theta = \theta_0) \approx P(\chi_1^2 \geq d) \tag{13.3.3}$$

on riittävän hyvä. Likiarvo saadaan siis χ^2-jakauman avulla, koska suurilla n:n arvoilla W noudattaa likimain χ^2-jakaumaa vapausastein 1.

Esimerkki 13.6 Tehdään n riippumatonta Bernoullin koetta, joissa onnistumistodennäköisyys on θ. Onnistumisten lukumäärä X noudattaa binomijakaumaa $\mathrm{Bin}(n, \theta)$. Testataan hypoteesia H_0: $\theta = \theta_0$, missä θ_0 on jokin numeerinen arvo.

Binomijakauman logaritmoitu uskottavuusfunktio on

$$l(\theta) = x \log\theta + (n - x)\log(1 - \theta),$$

missä $0 < \theta < 1$. Silloin

$$l(\hat{\theta}) = x \log \frac{x}{n} + (n - x) \log\left(1 - \frac{x}{n}\right),$$

missä $\hat{\theta} = x/n$. Jos H_0 on tosi, silloin logaritmoidun uskottavuusfunktion arvo on

$$l(\theta_0) = x \log \theta_0 + (n - x) \log(1 - \theta_0).$$

Jos esimerkiksi $\theta_0 = 0.5$, silloin $l(0.5) = n \log(0.5)$. Uskottavuustestisuure hypoteesin H_0: $\theta = \theta_0$ testaamiseksi on

$$\begin{aligned} W &= 2[l(\hat{\theta}) - l(\theta_0)] = -2\, r(\theta_0) \\ &= 2\left[x \log \frac{x}{n\theta_0} + (n - x) \log \frac{n - x}{n(1 - \theta_0)}\right] \\ &= 2\left[x \log \frac{\hat{\theta}}{\theta_0} + (n - x) \log \frac{1 - \hat{\theta}}{1 - \theta_0}\right]. \end{aligned}$$

Jos n on suuri, W noudattaa likimain χ^2-jakaumaa vapausastein 1.

Olkoon esimerkiksi $n = 100$ ja $\theta_0 = \frac{1}{4}$. Silloin uskottavuustestisuure hypoteesin H_0: $\theta = \frac{1}{4}$ testaamiseksi on

$$W = 2\left[x \log \frac{x}{25} + (n - x) \log \frac{100 - x}{75}\right].$$

Jos havaitaan $X = 43$, on W:n havaittu arvo

$$d = 86 \cdot \log \frac{43}{25} + 114 \cdot \log \frac{57}{75} = 15.35.$$

Testisuureen arvoon liittyvä p-arvo on

$$\alpha_h = P\big(W \geq 15.35 \;\big|\; \theta = \tfrac{1}{4}\big) \approx P(\chi_1^2 \geq 15.35) \approx 0.$$

Jos $\theta = \frac{1}{4}$, on erittäin epätodennäköistä, että havaitaan arvo $X = 43$. Havainnot tukevat vahvasti käsitystä, että $\theta \neq \frac{1}{4}$.

Jos esimerkiksi $n = 10$ ja $\theta_0 = \frac{1}{4}$, niin χ^2-jakaumaan perustuva likiarvo ei ole enää kovin tarkka. Oletetaan, että havaittiin $X = 5$. Silloin W:n havaittu arvo on

$$d = 10 \cdot \log \frac{5}{2.5} + 10 \cdot \log \frac{5}{7.5} = 2.88$$

ja tarkka p-arvo saadaan binomijakauman $\mathrm{Bin}\big(10, \frac{1}{4}\big)$ avulla:

$$\alpha_h = P(W \geq 2.88) = P(X = 0) + P(X \geq 5) = 0.1344.$$

□

13.3.2 Useita parametreja

Oletetaan, että malli riippuu tuntemattomien parametrien vektorista $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Olkoon $\boldsymbol{\theta}_0$ tämän vektorin annettu arvo. Uskottavuussuure hypoteesin

$$H_0\colon \boldsymbol{\theta} = \boldsymbol{\theta}_0 \tag{13.3.4}$$

testaamiseksi on

$$W = 2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta}_0)] = -2\,r(\boldsymbol{\theta}_0),$$

missä $\hat{\boldsymbol{\theta}}$ on parametrivektorin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaatti ja $r(\boldsymbol{\theta})$ on parametrien θ_1, θ_2, ..., θ_k yhteisen normeeratun uskottavuusfunktion logaritmi. Voidaan osoittaa, että W noudattaa likimain χ^2-jakaumaa, jonka vapausasteiden lukumäärä on mallissa olevien toisistaan funktionaalisesti riippumattomien parametrien lukumäärä. Tästä seuraa, että

$$\alpha_h = P(W \geq d \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0) \approx P[\chi^2(k) \geq d],$$

missä k on mallin riippumattomien parametrien lukumäärä

Esimerkki 13.7 Eräässä pitkäaikaisessa sydäntauteja koskevassa tutkimuksessa seurattiin suurta joukkoa miehiä, joilla ei ollut mitään merkkiä aikaisemmista sydänongelmista. Seurantajakson aikana miehistä 63 kuoli yllättävään sydäninfarktiin. Seuraavassa taulukossa kuolemat on kirjattu viikon päivän mukaan.

Viikon päivä	ma	ti	ke	to	pe	la	su
Kuolemien lkm	22	7	6	13	5	4	6

Testataan hypoteesi, että kuoleman todennäköisyys on sama kaikkina päivinä. Silloin testattava hypoteesi on

$$H_0\colon \theta_i = \frac{1}{7}, \qquad i = 1, \ldots, 7,$$

missä θ_i on kuoleman todennäköisyys viikon i. päivänä. Mallissa on 6 vapaata parametria, sillä parametreja $\theta_1, \ldots, \theta_7$ sitoo rajoite $\sum_{i=1}^{7} \theta_i = 1$. Olkoon f_i viikon i. päivänä kuolleiden lukumäärä. Koska $\hat{\theta}_i = f_i/63$, niin

$$l(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{7} f_i \log \hat{\theta}_i = -110.995.$$

Nollahypoteesin vallitessa

$$l\left(\tfrac{1}{7}\right) = \sum_{i=1}^{7} f_i \log\left(\tfrac{1}{7}\right) = -63 \log 7 = -122.592.$$

Silloin uskottavuustestisuureen arvo on

$$W = 2[l(\hat{\boldsymbol{\theta}}) - l\left(\tfrac{1}{7}\right)] = 23.274.$$

Testisuureeseen liittyvä p-arvo on

$$\alpha_h \approx P(\chi_6^2 \geq 23.274) = 0.0007.$$

Havainnot eivät tue hypoteesia H_0. □

13.4 Uskottavuusfunktion avulla konstruoituja testisuureita

Alaluvussa 13.3.1 esitettiin Wilksin *uskottavuustestisuure* $W = -2\log\frac{L(\theta;x)}{L(\hat{\theta};x)}$. Seuraavassa esitetään kaksi muuta keskeistä uskottavuusfunktioon perustuvaa testisuuretta. Nämä ovat *Waldin testisuure* ja *Raon pistetestisuure.*

Tuloksen (11.2.3) nojalla suurilla n:n arvoilla likimain

$$\hat{\theta} \sim \mathrm{N}[\theta, \mathcal{I}(\hat{\theta})^{-1}],$$

ja vastaavasti likimain

$$(\hat{\theta} - \theta)\sqrt{\mathcal{I}(\hat{\theta})} \sim \mathrm{N}(0,1).$$

ja

$$(\hat{\theta} - \theta)^2\,\mathcal{I}(\hat{\theta}) \sim \chi^2(1).$$

Kun testataan hypoteesia H_0: $\theta = \theta_0$, niin H_0:n vallitessa testisuure

$$\chi^2_W = (\hat{\theta} - \theta_0)^2\,\mathcal{I}(\hat{\theta}) \tag{13.4.1}$$

noudattaa likimain χ^2-jakaumaa vapausastein 1.

Vastaava tulos pätee myös vektoriarvoisille parametreille. Jos $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\mathsf{T}$ ja $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)^\mathsf{T}$, niin $\hat{\boldsymbol{\theta}}$ noudattaa suurilla n:n arvoilla likimain k:n muuttujan normaalijakaumaa $\mathrm{N}_k[\boldsymbol{\theta}, \mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}]$, missä k on vapaiden parametrien lukumäärä ja $\mathcal{I}(\hat{\boldsymbol{\theta}})$ on Fisherin informaatiomatriisin (odotettu informaatio) estimaatti. Kun H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, niin likimain

$$\chi^2_W = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\mathsf{T}\,\mathcal{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim \chi^2(k). \tag{13.4.2}$$

Testisuureita (13.4.1) ja (13.4.2) kutsutaan Waldin testisuureiksi.

Pistesuure $S(\theta) = l'(\theta; X_1, \ldots, X_n)$ noudattaa likimain normaalijakaumaa $\mathrm{N}[0, \mathcal{I}(\theta)]$, kun n on suuri. Silloin likimain

$$\frac{S(\theta)}{\sqrt{\mathcal{I}(\theta)}} \sim \mathrm{N}(0,1) \qquad \text{ja vastaavasti} \qquad \frac{S(\theta)^2}{\mathcal{I}(\theta)} \sim \chi^2(1).$$

Kun testataan hypoteesia H_0: $\theta = \theta_0$, niin H_0:n vallitessa testisuure

$$U = \frac{S(\theta_0)^2}{\mathcal{I}(\theta_0)} \tag{13.4.3}$$

noudattaa myös likimain χ^2-jakaumaa vapausastein 1.

Kun $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\mathsf{T}$, niin suurilla n:n arvoilla likimain

$$S(\boldsymbol{\theta})^\mathsf{T}\,\mathcal{I}(\boldsymbol{\theta})^{-1}S(\boldsymbol{\theta}) \sim \chi^2(k),$$

missä $S(\boldsymbol{\theta})$ on pistevektori ja $\mathcal{I}(\boldsymbol{\theta})^{-1}$ on informaatiomatriisin käänteismatriisi. Kun testataan hypoteesia H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, niin H_0:n vallitessa likimain

$$U = S(\boldsymbol{\theta}_0)^{\mathsf{T}} \mathcal{I}(\boldsymbol{\theta}_0)^{-1} S(\boldsymbol{\theta}_0) \sim \chi^2(k). \tag{13.4.4}$$

Testisuuretta (13.4.3) sanotaan Raon pistetestisuureeksi. Jos hypoteesi on yhdistetty ja testisuure (13.4.4) riippuu tuntemattomasta parametrista, ne estimoidaan H_0:n vallitessa. Raon pistetestisuureella, Wilksin uskottavuustestisuureella ja Waldin testisuureella on sama asymptoottinen jakauma.

Esimerkki 13.8 Olkoon $X_1, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\theta, \sigma^2)$, missä σ^2 on tunnettu. Testataan hypoteesia H_0: $\theta = \theta_0$, missä θ_0 on jokin annettu θ:n arvo. Silloin $\hat{\theta} = \overline{X}$ ja $\mathcal{I}(\theta) = n/\sigma^2$, joten

$$\chi^2_W = (\hat{\theta} - \theta_0)^2 \mathcal{I}(\hat{\theta}) = \frac{n}{\sigma^2}(\overline{X} - \theta_0)^2.$$

Huomaa, että $\mathcal{I}(\hat{\theta}) = \mathcal{I}(\theta)$, koska $\mathcal{I}(\theta)$ ei riipu parametrin arvosta. Koska $(\overline{X} - \theta_0)/(\sigma/\sqrt{n}) \sim \mathrm{N}(0,1)$, niin

$$\chi^2_W = \left(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1).$$

Tässä tapauksessa Waldin testisuure noudattaa täsmällisesti χ^2-jakaumaa.

Raon pistetestisuure on

$$U = \frac{S(\theta_0)^2}{\mathcal{I}(\theta_0)} = \frac{\sigma^2}{n}\left[\frac{n(\overline{X} - \theta_0)}{\sigma^2}\right]^2 = \frac{n}{\sigma^2}(\overline{X} - \theta_0)^2,$$

koska $U(\theta_0) = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \theta_0) = \frac{n}{\sigma^2}(\overline{X} - \theta_0)$.

Uskottavuustestisuure on

$$\begin{aligned} W &= 2[l(\hat{\theta}) - l(\theta_0)] \\ &= -\frac{1}{\sigma^2}\sum (X_i - \overline{X})^2 + \frac{1}{\sigma^2}\sum (X_i - \theta_0)^2 = \frac{n}{\sigma^2}(\overline{X} - \theta_0)^2. \end{aligned}$$

Tässä tapauksessa siis kaikki kolme testisuuretta W, U ja D ovat identtiset. □

13.5 Uskottavuussuhdetestit: Yhdistetyt hypoteesit

Oletetaan, että todennäköisyysmalli riippuu tuntemattomista parametreista $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Hypoteesi H_0 koskee nyt parametrivektorin $\boldsymbol{\theta}$ alkioita. Hypoteesi H_0 määrittää mallin, joka saadaan asettamalla joitain rajoitteita alkuperäiseen malliin. Hypoteesin H_0 määrittämä malli on alkuperäistä mallia yksinkertaisempi ja siinä on $q < k$ funktionaalisesti riippumatonta parametria.

Yksinkertainen hypoteesi määrittää kaikille parametreille arvot siten, että H_0:n vallitessa parametrien arvot on täysin määrätty eli $q = 0$. Yhdistetty hypoteesi ei poista kaikkia tuntemattomia parametreja ja silloin $q > 0$.

Olkoon $l(\boldsymbol{\theta})$ parametrivektorin $\boldsymbol{\theta}$ logaritmoitu uskottavuusfunktio perusmallissa. Silloin $l(\hat{\boldsymbol{\theta}}) \geq l(\boldsymbol{\theta})$ kaikilla $\boldsymbol{\theta}$:n arvoilla, kun $\hat{\boldsymbol{\theta}}$ on $\boldsymbol{\theta}$:n suurimman uskottavuuden estimaatti.

Olkoon $\tilde{\boldsymbol{\theta}}$ parametrivektorin $\boldsymbol{\theta}$ suurimman uskottavuuden estimaatti, kun oletetaan H_0 todeksi. Koska $l(\hat{\boldsymbol{\theta}}) \geq l(\boldsymbol{\theta})$ kaikilla $\boldsymbol{\theta}$:n arvoilla, niin $l(\hat{\boldsymbol{\theta}}) \geq l(\tilde{\boldsymbol{\theta}})$. Rajoitteen H_0 vallitessa laskettu maksimiarvo $l(\tilde{\boldsymbol{\theta}})$ ei voi olla suurempi kuin ilman rajoitetta laskettu maksimiarvo $l(\hat{\boldsymbol{\theta}})$.

Uskottavuussuhdetestisuure hypoteesin H_0 testaamiseksi määritellään näiden kahden maksimin erotuksen avulla:

$$W = 2[l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}})]. \tag{13.5.1}$$

Testisuureen W nimi johtuu siitä, että se voidaan lausua uskottavuussuhteen $L(\hat{\boldsymbol{\theta}})/L(\tilde{\boldsymbol{\theta}})$ funktiona seuraavasti:

$$W = 2 \log \frac{L(\hat{\boldsymbol{\theta}})}{L(\tilde{\boldsymbol{\theta}})}.$$

Koska $l(\hat{\boldsymbol{\theta}}) \geq l(\tilde{\boldsymbol{\theta}})$, niin W on epänegatiivinen. Kun hypoteesi H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on yksinkertainen, niin vektori $\boldsymbol{\theta}_0$ on jokin annettu numeerinen vektori. Silloin H_0:n vallitessa vain arvo $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on mahdollinen, joten $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ ja funktion $l(\boldsymbol{\theta})$ maksimi yksinkertaisen hypoteesin tapauksessa on $l(\boldsymbol{\theta}_0)$.

13.5.1 p-arvon määrittäminen

Okoon $W = d$ havaittu uskottavuustestisuureen arvo. Silloin testisuureen arvoon liittyvä p-arvo on

$$\alpha_h = P(d \geq d \mid H_0 \text{ tosi}).$$

Useimmiten α_h:n tarkan arvon laskeminen on hankalaa, mutta χ^2-jakauman perusteella saadaan tyydyttävä likiarvo. Melko yleisten oletusten vallitessa (vrt. alaluku **??**) *W noudattaa χ^2-jakaumaa vapausastein $k - q$, kun H_0 on tosi.* Tämän likiarvon perusteella

$$\alpha_h \approx P(\chi^2_{k-q} \geq d).$$

Huomattakoon, että χ^2-jakauman vapausasteet $k - q$ saadaan perusmallin parametrien lukumäärän k ja H_0:n määrittämän mallin parametrien lukumäärän q erotuksena.

13.5.2 Kaksi parametria, joista toista testataan

Oletetaan, että $\boldsymbol{\theta} = (\theta_1, \theta_2)$, joten $k = 2$. Testataan hypoteesia

$$H_0\colon \theta_2 = a,$$

missä a on annettu numeerinen arvo. Hypoteesin H_0 vallitessa mallissa on yksi tuntematon parametri θ_1, joten $q = 1$.

Olkoon $l(\theta_1, \theta_2)$ parametrien θ_1 ja θ_2 logaritmoitu uskottavuusfunktio. Tavallisesti parametrien θ_1 ja θ_2 suurimman uskottavuuden estimaatit $\hat{\theta}_1$ ja $\hat{\theta}_2$ saadaan ratkaisemalla samanaikaisesti uskottavuusyhtälöt

$$s_1(\theta_1, \theta_2) = 0 \qquad \text{ja} \qquad s_2(\theta_1, \theta_2) = 0,$$

missä $s_1(\theta_1, \theta_2) = \frac{\partial\, l(\theta_1, \theta_2)}{\partial \theta_1}$ ja $s_2(\theta_1, \theta_2) = \frac{\partial\, l(\theta_1, \theta_2)}{\partial \theta_2}$. Kun H_0:n mukaisesti $\theta_2 = a$, saadaan θ_1:n ehdollinen suurimman uskottavuuden estimaatti $\hat{\theta}_1(a)$ ratkaisemalla yhtälö

$$s_1(\theta_1, a) = \frac{\partial\, l(\theta_1, a)}{\partial \theta_1} = 0.$$

Logaritmoidun uskottavuusfunktion $l(\theta_1, \theta_2)$ maksimi on $l(\hat{\theta}_1, \hat{\theta}_2)$. Hypoteesin H_0 vallitessa logaritmoidun uskottavuusfunktion maksimi on $l[\hat{\theta}_1(a), a]$. Siksi uskottavuustestisuure hypoteesin H_0: $\theta_2 = a$ testaamiseksi on

$$W = 2\big[l(\hat{\theta}_1, \hat{\theta}_2) - l\big(\hat{\theta}_1(a), a\big)\big].$$

Uskottavuusfunktio $l(\theta_1, \theta_2)$ maksimoidaan ensin ilman rajoitteita ja sitten rajoitteen H_0: $\theta_2 = a$ vallitessa ja lasketaan maksimien erotus. Kun H_0: $\theta_2 = a$ on tosi, niin W:n jakauman likiarvona voidaan käyttää χ^2-jakaumaa vapausastein $k - q = 2 - 1 = 1$. Vastaavasti p-arvon likiarvo on

$$\alpha_h = P(W \geq d \mid \theta_2 = a) \approx P(\chi^2_1 \geq d).$$

Esimerkki 13.9 Olkoon $x_1, \ldots, x_n$ otos Weibullin jakaumasta, jonka tiheysfunktio on

$$f(x) = \alpha\beta x^{\beta-1} \exp(-\alpha x^{\beta}), \qquad 0 < x < \infty, \tag{13.5.2}$$

missä $\alpha > 0$ ja $\beta > 0$ ovat tuntemattomia parametreja. Testataan esimerkiksi iskunvaimentimien kestoa kokeilemalla, montako standardivoimakkuudella annettua iskua ne kestävät tietyissä vakio-olosuhteissa. Tuote-erästä valittiin kokeeseen 25 iskunvaimenninta.

Kun $\beta = 1$, saadaan Weibullin jakaumasta (13.5.2) erikoistapauksena eksponenttijakauma. Testataan siksi hypoteesi H_0: $\beta = 1$. Mallin (13.5.2) logaritmoitu uskottavuusfunktio on

$$l(\alpha, \beta) = n \log \alpha + n \log \beta + (\beta - 1) \sum \log x_i - \alpha \sum x_i^{\beta}.$$

Pistefunktiot ovat

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - \sum x_i^\beta,$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} + \sum \log x_i - \alpha \sum x_i^\beta \cdot \log x_i.$$

Yhtälön $\frac{\partial l}{\partial \alpha} = 0$ ratkaisu on $\hat{\alpha}(\beta) = n \big/ \sum x_i^\beta$. Sijoitetaan tämä ratkaisu yhtälöön $\frac{\partial l}{\partial \beta} = s_2(\alpha, \beta) = 0$ ja ratkaistaan β. Saadaan

$$s_2[\hat{\alpha}(\beta), \beta] = \frac{n}{\beta} + \sum \log x_i - \frac{n \sum x_i^\beta \log x_i}{\sum x_i^\beta}.$$

Yhtälö $s_2[\hat{\alpha}(\beta), \beta] = 0$ voidaan ratkaista numeerisesti esimerkiksi Newtonin menetelmällä.

Aineistosta saatiin suurimman uskottavuuden estimaatit

$$\hat{\alpha} = 9.515 \cdot 10^{-5} \qquad \text{ja} \qquad \hat{\beta} = 2.1021.$$

Uskottavuusfunktion maksimi on

$$l(\hat{\alpha}, \hat{\beta}) = -113.691.$$

Kun H_0: $\beta = 1$ on tosi, niin α:n suurimman uskottavuuden estimaatti on

$$\hat{\alpha}(1) = \frac{n}{\sum x_i} = \frac{25}{1805} = 0.01385$$

ja uskottavuusfunktion maksimi

$$l\big(\hat{\alpha}(1), 1\big) = -121.433.$$

Uskottavuustestisuureen arvo testattaessa hypoteesia H_0: $\beta = 1$ on

$$W = 2\big[l(\hat{\alpha}, \hat{\beta}) - l\big(\hat{\alpha}(1), 1\big)\big] = 2(-113.691 + 121.433) = 15.48.$$

Testisuureen arvoon liittyvä p-arvo on

$$\alpha_h \approx P(\chi_1^2 \geq 15.48) < 0.001.$$

Havainnot tukevat vahvasti käsitystä, että H_0: $\beta = 1$ ei pidä paikkaansa. □

13.5.3 Homogeenisuuden testaus

Tehdään k riippumatonta koetta. Oletetaan aluksi, että jokaisen kokeen tuloksista estimoidaan eri parametri θ_i, $i = 1, \ldots, k$. Olkoot $l_i(\theta)$ kokeisiin $i = 1, \ldots, k$ liittyvät logaritmoidut uskottavuusfunktiot ja $\hat{\theta}_i$ vastaavat suurimman uskottavuuden estimaatit. Yhdistettyyn kokeeseen liittyvä uskottavuusfunktio on

$$l(\theta_1, \ldots, \theta_k) = l_1(\theta_1) + \cdots + l_k(\theta_k)$$

ja funktion maksimi on $\sum_{i=1}^{k} l(\hat{\theta}_i)$.

Testataan nyt hypoteesia

$$H_0\colon \theta_1 = \cdots = \theta_k = \theta. \tag{13.5.3}$$

Hypoteesin H_0 vallitessa kaikissa kokeissa parametrin arvo on sama ja silloin logaritmoitu uskottavuusfunktio on

$$l(\theta_1, \ldots, \theta_k) = \sum l_i(\theta), \tag{13.5.4}$$

missä yhteistä parametrin arvoa on merkitty θ:lla. Merkitään yhdistetystä kokeesta estimoitua suurimman uskottavuuden estimaattia $\tilde{\theta}$:lla, jolloin funktion (13.5.4) maksimi on $\sum l_i(\tilde{\theta})$. Uskottavuustestisuure hypoteesin (13.5.3) testaamiseksi on

$$W = 2\sum_{i=1}^{k}\Big[l_i(\hat{\theta}_i) - l_i(\tilde{\theta})\Big] = -2\sum_{i=1}^{k} r_i(\tilde{\theta}),$$

missä $r_i(\theta_i) = l_i(\theta_i) - l_i(\hat{\theta}_i)$ on i. kokeeseen liittyvä logaritmoitu normitettu uskottavuusfunktio.

Jos W on suuri, niin silloin ei ole olemassa sellaista parametrien arvoa, joka olisi uskottava kaikissa kokeissa. Hypoteesin testauksessa on $k-1$ vapausastetta, koska H_0 pudottaa tuntemattomien parametrien lukumäärän k:sta parametrista yhteen parametriin. Jos havaitaan $W = d$, niin

$$\alpha_h \approx P(\chi^2_{k-1} \geq d).$$

Esimerkki 13.10 Olkoon $X_1, \ldots, X_n$ otos eksponenttijakaumasta $\mathrm{Exp}(\theta_1)$ ja $Y_1 \ldots, Y_m$ otos eksponenttijakaumasta $\mathrm{Exp}(\theta_2)$. Silloin

$$f(x_1, \ldots, x_n; \theta_1) = \prod_{i=1}^{n} \frac{1}{\theta_1}\,\mathrm{e}^{-x_i/\theta_1} = \frac{1}{\theta_1^n}\,\mathrm{e}^{-\sum x_i/\theta_1}$$

ja

$$f(y_1, \ldots, y_m; \theta_2) = \prod_{j=1}^{m} \frac{1}{\theta_2}\,\mathrm{e}^{-y_j/\theta_2} = \frac{1}{\theta_2^m}\,\mathrm{e}^{-\sum y_j/\theta_2}.$$

Parametrin θ_1 logaritmoitu uskottavuusfunktio on

$$l_1(\theta_1) = -n\log\theta_1 - n\overline{X}/\theta_1$$

ja θ_2:n logaritmoitu uskottavuusfunktio on

$$l_2(\theta_2) = -m\log\theta_2 - m\overline{Y}/\theta_2,$$

missä $n\overline{X} = \sum_{i=1}^{n} X_i$ ja $m\overline{Y} = \sum_{j=1}^{m} Y_j$. Koska $\hat{\theta}_1 = \overline{X}$ ja $\hat{\theta}_2 = \overline{Y}$, niin

$$l_1(\hat{\theta}_1) = -n\log\overline{X} - n$$

ja

$$l_2(\hat{\theta}_2) = -m \log \overline{Y} - m.$$

Testataan hypoteesia $H_0 : \theta_1 = \theta_2$. Jos H_0 on tosi, niin $\theta_1 = \theta_2 = \theta$ ja silloin

$$X_1, \ldots, X_n, Y_1, \ldots, Y_m \sim \mathrm{Exp}(\theta).$$

Parametrin θ logaritmoitu uskottavuusfunktio on

$$l(\theta) = -(n+m)\log\theta - \frac{\sum X_i + \sum Y_j}{\theta}$$

ja θ:n suurimman uskottavuuden estimaatti on yhdistetyn otoksen keskiarvo

$$\tilde{\theta} = \frac{\sum X_i + \sum Y_j}{n+m} = \frac{n\overline{X} + m\overline{Y}}{n+m}.$$

Silloin

$$l(\tilde{\theta}) = -(n+m)\log\tilde{\theta} - (n+m)$$

ja

$$\begin{aligned} W &= 2[l_1(\hat{\theta}_1) + l_2(\hat{\theta}_2) - l(\tilde{\theta})] \\ &= 2[-n\log\overline{X} - m\log\overline{Y} - (n+m) + (n+m)\log\tilde{\theta} + (n+m)] \\ &= 2\left[n\log\frac{\tilde{\theta}}{\overline{X}} + m\log\frac{\tilde{\theta}}{\overline{Y}}\right], \end{aligned}$$

joka riippuu vain otoskeskiarvoista $\overline{X}$, $\overline{Y}$ ja otosten koosta n ja m. □

Esimerkki 13.11 Testataan, onko erään järven vedessä tiettyä bakteeria. Ei ole mahdollista laskea näytteessä olevien bakteerien määrää, vaan voidaan ainoastaan todeta, onko näyteputkessa bakteereita vai ei. Negatiivinen tulos osoittaa, että näyteputkessa ei ole bakteereita. Positiivinen tulos tarkoittaa, että näytteessä on ainakin yksi bakteeri.

Bakteerien lukumäärä X tilavuudessa V vettä noudattaa Poissonin jakaumaa $\mathrm{Poi}(\mu V)$:

$$f(x) = \frac{(\mu V)^x \mathrm{e}^{-\mu V}}{x!}, \qquad x = 0, 1, \ldots.$$

Negatiivisen reaktion todennäköisyys on

$$\pi = f(0) = \mathrm{e}^{-\mu V}$$

ja positiivisen reaktion todennäköisyys on

$$1 - \pi = 1 - \mathrm{e}^{-\mu V}.$$

Kun testataan n vesinäytettä, noudattaa negatiivisten näytteiden lukumäärä Y binomijakaumaa $\mathrm{Bin}(n, \pi)$:

$$P(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}.$$

Kun vesinäytteitä otetaan, on määritettävä yhteen näytteeseen tarvittavan vesimäärän tilavuus. Jos V on liian suuri, kaikkiin näytteisiin tulee bakteereita ja kaikki näytteet ovat positiivisia. Jos taas V on liian pieni, on vaara, että kaikki näytteet ovat negatiivisia. Eräs tapa suojautua tätä ongelmaa vastaan on valmistaa erikokoisia näytteitä.

Tehtiin kaksi riippumatonta koetta. Ensimmäisessä kokeessa oli 40 näyteputkea, joiden tilavuus $V = 10\,\mathrm{ml}$. Toisessa kokeessa tarkistettiin 40 vesinäytettä, joiden tilavuus oli $1\,\mathrm{ml}$. Kun $V = 10\,\mathrm{ml}$, saatiin 28 negatiivista ja 12 positiivista tulosta. Kun näytteen koko $V = 1\,\mathrm{ml}$, saatiin 37 negatiivista ja 3 positiivista tulosta.

Ensimmäisessä kokeessa uskottavuusfunktio on

$$L_1(\mu_1) = \pi_1^{28}(1 - \pi_1)^{12},$$

missä $\pi_1 = \mathrm{e}^{-10\mu_1}$. Parametrin μ_1 suurimman uskottavuuden estimaatti on $\hat{\mu}_1 = 0.0357$. Toisesta kokeesta saadaan uskottavuusfunktioksi

$$L_2(\mu_2) = \pi_2^{37}(1 - \pi_2)^3,$$

missä $\pi_2 = \mathrm{e}^{-\mu_2}$. Parametrin μ_2 suurimman uskottavuuden estimaatti on $\hat{\mu}_2 = 0.0780$. Kokeisiin liittyvät logaritmoidut uskottavuusfunktiot ovat vastaavasti

$$l_1(\mu_1) = 28 \log \pi_1 + 12 \log(1 - \pi_1)$$

ja

$$l_2(\mu_2) = 37 \log \pi_2 + 3 \log(1 - \pi_2).$$

Jos oletetaan, että bakteerien määrä millilitrassa on molemmissa kokeissa keskimäärin sama, voidaan kokeet yhdistää. Kaikkiin 80 näytteeseen perustuvan yhdistetyn kokeen parametrin μ logaritmoitu uskottavuusfunktio on

$$l(\mu) = l_1(\mu) + l_2(\mu).$$

Tästä funktiosta maksimoimalla saatu suurimman uskottavuuden estimaatti $\tilde{\mu} = 0.04005$. Ensimmäisestä kokeesta saadaan logaritmoitu normitettu uskottavuusfunktio

$$r_1(\mu) = -280\mu + 12 \log(1 - \mathrm{e}^{-10\mu}) + 24.43$$

ja toisesta kokeesta

$$r_2(\mu) = -37\mu + 3 \log(1 - \mathrm{e}^{-\mu}) + 10.66.$$

Kun testataan hypoteesia

$$H_0\colon \mu_1 = \mu_2,$$

on uskottavuustestisuureen arvo

$$W = 2[l_1(\hat{\mu}_1) + l_2(\hat{\mu}_2) - l_1(\tilde{\mu}) - l_2(\tilde{\mu})] = -2[r_1(\tilde{\mu}) + r_2(\tilde{\mu})] = 1.24.$$

Hypoteesin testauksessa on yksi vapausaste, joten

$$\alpha_h \approx P(\chi_1^2 \geq 1.24) > 0.25.$$

Havainnot ovat sopusoinnussa H_0:n kanssa. □

13.5.4 Binomitodennäköisyyksien testaaminen

Oletetaan, että verrataan erilaisten hoitojen vaikutusta. Olkoon vertailtavana k erilaista hoitomenetelmää. Ensimmäistä hoitoa annettiin n_1:lle potilaalle, joista Y_1 parani. Vastaavasti i. hoito annettiin n_i:lle potilaalle, joista Y_i parani. Tällaisen hoitokokeen tulokset ovat Taulukossa 13.1 esitettyä muotoa.

Taulukko 13.1. Hoitokokeiden tulokset parantuneiden lukumäärän mukaan.

Käsittely	1	2	...	k
Parantuneiden lkm	Y_1	Y_2	...	Y_k
Epäonnistumisten lkm	$n_1 - Y_1$	$n_2 - Y_2$	...	$n_k - Y_k$
Potilaiden lkm	n_1	n_2	...	n_k

Oletamme, että parantuneiden lukumäärä Y_i hoidossa i noudattaa binomijakaumaa $\mathrm{Bin}(n_i, \pi_i)$ ja että lukumäärät $Y_1, \ldots, Y_k$ ovat toisistaan riippumattomat. Hoitojen tehoa voidaan vertailla onnistumistodennäköisyyksien $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ avulla. Perusmallissa on k tuntematonta parametria ja logaritmoitu uskottavuusfunktio on

$$l(\boldsymbol{\pi}) = \sum_{i=1}^{k} y_i \log \pi_i + \sum_{i=1}^{k} (n_i - y_i) \log(1 - \pi_i).$$

Parametrin π_i suurimman uskottavuuden estimaatti on $\hat{\pi}_i = y_i/n_i$ ja $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_k)$. Logaritmoidun uskottavuusfunktion maksimi on

$$l(\hat{\boldsymbol{\pi}}) = \sum y_i \log \frac{y_i}{n_i} + \sum (n_i - y_i) \log\Bigl(1 - \frac{y_i}{n_i}\Bigr).$$

Testataan nyt hypoteesi, että hoidoilla ei ole eroa:

$$H_0\colon \pi_1 = \pi_2 = \cdots = \pi_k = \pi. \tag{13.5.5}$$

Hypoteesin H_0 mukaisessa mallissa on yksi tuntematon parametri, koska onnistumistodennäköisyyksien yhteinen arvo on tuntematon. Kun H_0 on tosi on logaritmoitu uskottavuusfunktio

$$l(\pi) = \sum y_i \log \pi + \sum (n_i - y_i) \log(1 - \pi), \tag{13.5.6}$$

missä π on hoitojen yhteinen onnistumistodennäköisyys. Logaritmoidusta uskottavuusfunktiosta (13.5.6) määritetty suurimman uskottavuuden estimaatti on $\tilde{\pi} = y/n$ ja $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}, \ldots, \tilde{\pi})$, missä $y = \sum_{i=1}^{k} y_i$ ja $n = \sum_{i=1}^{k} n_i$. Logaritmoidun uskottavuusfunktion (13.5.6) maksimi on

$$l(\tilde{\boldsymbol{\pi}}) = \sum_{i=1}^{k} y_i \log \tilde{\pi} + \sum_{i=1}^{k} (n_i - y_i) \log(1 - \tilde{\pi}).$$

Uskottavuustestisuure hypoteesin (13.5.5) testaamiseksi on

$$W = 2[l(\hat{\boldsymbol{\pi}}) - l(\tilde{\boldsymbol{\pi}})] = 2\left[\sum_{i=1}^{k} y_i \log \frac{y_i}{n_i \tilde{\pi}} + \sum_{i=1}^{k} (n_i - y_i) \log \frac{n_i - y_i}{n_i(1 - \tilde{\pi})}\right]. \tag{13.5.7}$$

Huomattakoon, että onnistumisten lukumäärien odotusarvot ovat $n_i\pi$ ja epäonnistumisten lukumäärien odotusarvot $n_i(1-\pi)$, jos H_0 on tosi. Vastaavasti $n_i\tilde{\pi}$ ja $n_i(1-\tilde{\pi})$ ovat näiden odotettujen frekvenssien estimaatit ja $y_1, \ldots, y_k$ ovat havaitut frekvenssit. Testisuure (13.5.7) vertailee siis havaittuja ja H_0:n vallitessa estimoituja frekvenssejä yli kaikkien luokkien ($2k$ kappaletta).

Jos hoidot ovat lääkkeen eri annosmääriä a_1, a_2, $\ldots$, a_k, niin voitaisiin testata hypoteesia

$$H\colon \pi_i = 1 - \frac{1}{1 + \mathrm{e}^{\alpha + \beta a_i}}, \qquad i = 1, \ldots, k. \tag{13.5.8}$$

Tässää hypoteesissa oletetaan, että paranemistodennäköisyys riippuu annoksesta logistisen mallin mukaan. Hypoteesin (13.5.8) spesifioimassa mallissa on kaksi tuntematonta parametria. Kun hypoteesi (13.5.8) oletetaan oikeaksi ja kirjoitetaan logaritmoitu uskottavuusfunktio parametrien α ja β funktiona, voidaan määrittää niiden suurimman uskottavuuden estimaatit $\tilde{\alpha}$ ja $\tilde{\beta}$. Niiden avulla voidaan laskea sitten estimaatit

$$\tilde{\pi}_i = 1 - \frac{1}{1 + \mathrm{e}^{\tilde{\alpha} + \tilde{\beta} a_i}}, \qquad i = 1, \ldots, k.$$

Koska $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_k)$, niin uskottavuustestisuure on

$$W = 2\left[\sum y_i \log \frac{y_i}{n_i \tilde{\pi}_i} + \sum (n_i - y_i) \log \frac{n_i - y_i}{n_i(1 - \tilde{\pi}_i)}\right].$$

Kun havaitaan $W = d$, niin

$$\alpha_h \approx P(\chi^2_{k-q} \geq d).$$

Hypoteesin (13.5.5) tapauksessa $q = 1$ ja hypoteesin (13.5.8) tapauksessa $q = 2$. Likiarvon voidaan oletaa olevan kohtuullisen tarkka, mikäli odotettujen frekvenssien estimaatit ovat kohtuullisen suuria.

Esimerkki 13.12 Tutkittiin erään ruuan lisäaineen mahdollisesti syöpää aiheuttavaa vaikutusta rotilla siten, että 44 rotalle annettiin ainetta pieni annos ja 44 rotalle suuri annos. Myöhemmin tutkittiin mahdollisesti kehittyneet kasvaimet. Tulokset ovat Taulukossa 13.2.

Taulukko 13.2. Kasvainten lukumäärät koeaineistossa rotilla lisäaineen annoksen mukaan.

Käsittely	Pieni annos	Suuri annos
Kasvain	4	14
Ei kasvainta	40	30
Yhteensä	44	44

Olkoot Y_1 ja Y_2 niiden rottien lukumäärät, joilla oli kasvain. Nyt $Y_i \sim \mathrm{Bin}(n_i, \pi_i)$, missä $n_1 = n_2 = 44$ ja π_1 on todennäköisyys saada kasvain pienellä annoksella ja π_2 todennäköisyys saada kasvain suurella annoksella. Testataan hypoteesia

$$H_0\colon \pi_1 = \pi_2.$$

H_0:n vallitessa kasvaintodennäköisyyden suurimman uskottavuuden estimaatti on

$$\tilde{\pi} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{4 + 14}{44 + 44} = \frac{9}{44}.$$

Odotettujen frekvenssien estimaatit ovat

$$n_1\tilde{\pi} = n_2\tilde{\pi} = 9 \qquad \text{ja} \qquad n_1(1 - \tilde{\pi}) = n_2(1 - \tilde{\pi}) = 35.$$

Lausekkeen (13.5.7) mukaan uskottavuussuhdetestisuureen arvo on

$$W = 2\big(4\log\tfrac{4}{9} + 14\log\tfrac{14}{9} + 40\log\tfrac{40}{35} + 30\log\tfrac{30}{35}\big) = 7.32.$$

Koska nyt

$$\alpha_h \approx P(\chi^2_{2-1} \geq 7.32) < 0.01,$$

koetulokset tukevat vahvasti käsitystä, että H_0: $\pi_1 = \pi_2$ ei pidä paikkaansa. □

13.5.5 Multinomitodennäköisyyksien testaaminen

Tarkastellaan koetta, jossa on k toisensa poissulkevaa tulosvaihtoehtoa T_1, …, T_k. Kun tehdään n riippumatonta toistoa, saadaan aineisto f_1, …, f_k, missä f_i on niiden kokeiden lukumäärä, joissa saadaan tulokseksi T_i.

Taulukossa 13.3 on vuoden aikana sattuneet polkupyöräonnettomuudet luokiteltuna viikonpäivän mukaan. Tässä esimerkissä $k = 7$ ja $n = \sum_{i=1}^{k} f_i = 174$. Frekvenssien todennäköisyydet saadaan multinomijakaumasta

$$p(f_1, \ldots, f_k) = \frac{n!}{f_1! f_2! \cdots f_k!}\pi_1^{f_1}\pi_2^{f_2}\cdots\pi_k^{f_k},$$

Taulukko 13.3. Polkupyöräonnettomuuksien lukumäärä eräässä kaupungissa vuonna 2003 viikon päivän mukaan.

Päivä	ma	ti	ke	to	pe	la	su	Yhteensä
Onnettomuuksien lkm	28	26	23	30	33	21	12	174

missä π_i on tulosvaihtoehdon T_i todennäköisyys. Parametrin $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ logaritmoitu uskottavuusfunktio on

$$l(\boldsymbol{\pi}) = \sum_{i=1}^{k} f_i \log \pi_i,$$

missä $\sum_{i=1}^{k} \pi_i = 1$. Parametrin π_i suurimman uskottavuuden estimaatti on $\hat{\pi}_i = f_i/n$. Siksi logaritmoidun uskottavuusfunktion maksimi on

$$l(\hat{\boldsymbol{\pi}}) = \sum_{i=1}^{k} f_i \log\left(\frac{f_i}{n}\right).$$

Hypoteesi määrittää parametreille (tai osalle niistä) numeeriset arvot tai esittää ne tuntemattomien parametrien funktiona. Merkitään H_0:n vallitessa laskettuja parametrien $\pi_1, \ldots, \pi_k$ suurimman uskottavuuden estimaatteja $\tilde{\pi}_1, \ldots, \tilde{\pi}_k$. Logaritmoidun uskottavuusfunktion maksimi H_0:n vallitessa on

$$l(\tilde{\boldsymbol{\pi}}) = \sum_{i=1}^{k} f_i \log \tilde{\pi}_j.$$

Edellä esitettyjen yleisten periaatteiden mukaan uskottavuustestisuure H_0:n testaamiseksi on muotoa

$$W = 2[l(\hat{\boldsymbol{\pi}}) - l(\tilde{\boldsymbol{\pi}})] = 2\sum_{i=1}^{k} f_i \log \frac{f_i}{e_i}, \tag{13.5.9}$$

missä $e_i = n\tilde{\pi}_i$ on odotetun frekvenssin estimaatti H_0:n vallitessa.

Koska $\sum_{i=1}^{k} \pi_i = 1$, niin perusmallissa on $k-1$ vapaata parametria. Taulukon 13.3 aineistolla voitaisiin esimerkiksi testata hypoteesia

$$H_0\colon \pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5 = \pi_6 = \pi_7.$$

H_0:n mukaan kaikkina viikon päivinä on sama onnettomuustodennäköisyys $\pi_i = \frac{1}{7}, i = 1, \ldots, 7$. Silloin siis todennäköisyydet on täysin määrätty, joten mallissa ei H_0:n vallitessa ole vapaita parametreja. Jos havaitaan $W = d$, niin

$$\alpha_h \approx P(\chi^2_{k-1} \geq d) = P(\chi^2_6 \geq d).$$

Ehkä realistisempi hypoteesi olisi olettaa, että viikon arkipäivinä ja viikonloppuna on erilainen onnettomuustodennäköisyys:

$$H\colon\ \pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5 = \pi_v \quad \text{ja} \quad \pi_6 = \pi_7 = \pi_s.$$

Silloin $5\,\pi_v + 2\,\pi_s = 1$, joten hypoteesin vallitessa mallissa on yksi vapaa parametri. Silloin siis $q = 1$ ja $k - 1 - q = 5$.

Luokilla, joissa $e_i \approx 0$ ja $f_i \geq 1$, on suuri vaikutus testisuureeseen W. Tavallisesti käytetään peukalosääntöä, että e_i:n tulisi olla vähintään 5, jotta χ^2-jakaumaan perustuva likiarvo olisi riittävän hyvä. Toinen tavanomainen multinomitodennäköisyyksien testaamiseen käytettävä testisuure on Pearsonin testisuure

$$X^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}. \tag{13.5.10}$$

Myös X^2 noudattaa asymptoottisesti χ^2-jakaumaa samoin vapausastein kuin W. Suureisiin (13.5.9) ja (13.5.10) perustuvia testejä sanotaan usein yhteensopivuustesteiksi.

13.5.6 Riippumattomuuden testaus kontingenssitaulukoissa

Oletetaan nyt, että tilastoyksiköt luokitellaan muuttujien X ja Y suhteen toisensa poissulkeviin luokkiin. Olkoon X esimerkiksi hiusten väri ja Y silmien väri. Hiukset luokitellaan vaaleisiin (Va) tummiin (T) ja punaisiin (P) sekä silmät sinisiin (S), ruskeisiin (R) ja vihreisiin (Vi). Jos valitaan n henkilön otos, saadaan Taulukon 13.4 kaltainen aineisto. Tällaista kahden muuttujan

Taulukko 13.4. Hiusten ja silmien värin välinen ristiintaulukko.

		Silmien väri			
		S	R	Vi	Yhteensä
Hiusten väri	Va	f_{11}	f_{12}	f_{13}	$f_{1.}$
	T	f_{21}	f_{22}	f_{23}	$f_{2.}$
	P	f_{31}	f_{32}	f_{33}	$f_{3.}$
Yhteensä		$f_{.1}$	$f_{.2}$	$f_{.3}$	n

suhteen esitettyä frekvenssitaulukkoa kutsutaan kontingenssitaulukoksi tai ristiintaulukoksi. Taulukossa 13.4 hiusten väriä kutsutaan rivimuuttujaksi ja simien väriä sarakemuuttujaksi.

Taulukossa 13.5 on kaikkiaan $k = IJ$ solua. Olkoon π_{ij} todennäköisyys, että havainto, osuu soluun (i, j). Kun koetta toistetaan n kertaa (tai tehdään n:n alkion satunnaisotos), noudattavat frekvenssit f_{ij} multinomijakaumaa. Parametrivektorin $\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \ldots, \pi_{IJ})$ logaritmoitu uskottavuusfunktio on

$$l(\boldsymbol{\pi}) = \sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij} \log \pi_{ij},$$

Taulukko 13.5. $I \times J$-kontingenssitaulukko.

	S_1	S_2	$\dots$	S_J	Yhteensä
R_1	f_{11}	f_{12}	$\dots$	f_{1J}	$f_{1.}$
R_2	f_{21}	f_{22}	$\dots$	f_{2J}	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
R_I	f_{I1}	f_{I2}	$\dots$	f_{IJ}	$f_{I.}$
Yhteensä	$f_{.1}$	$f_{.2}$	$\dots$	$f_{.J}$	n

missä $\sum\sum \pi_{ij} = 1$. Tässä tilanne on aivan samanlainen kuin multinomijakauman tapauksessa, paitsi että nyt käytetään kaksinkertaista indeksointia.

Todennäköisyyksiä π_{ij} koskevat hypoteesit H testaan uskottavuustestisuureella

$$W = 2\sum_{i=1}^{I}\sum_{j=1}^{J} f_{ij}\log\frac{f_{ij}}{e_{ij}},$$

missä $e_{ij} = n\tilde{\pi}_{ij}$ on luokan R_iS_j odotetun frekvenssin estimaatti ja $\tilde{\pi}_{ij}$ on solutodennäköisyyden π_{ij} suurimman uskottavuuden estimaatti H_0:n vallitessa.

Riippumattomuushypoteesi

Esimerkki 13.13 Usein ollaan kiinnostuneita siitä, onko rivi- ja sarakemuuttujien välillä riippuvuutta vai ovatko muuttujat riippumattomat. Ovatko esimerkiksi silmien ja hiusten väri toisistaan riippumattomat? Riippumattomuushypoteesi voidaan lausua seuraavasti:

(13.5.11) $$H_0\colon P(R_iS_j) = P(R_i)\,P(S_j) \qquad \text{kaikilla } i \text{ ja } j.$$

$P(R_iS_j) = \pi_{ij}$ on todennäköisyys, että tilastoyksikkö sattuu soluun (i,j). Vastaavasti

$$P(R_i) = \pi_{i1} + \cdots + \pi_{iJ} = \pi_{i.}$$

on todennäköisyys, että tilastoyksikkö sattuu rivimuuttujan i. luokkaan. Todennäköisyys, että tilastoyksikkö kuuluu sarakemuuttujan j. luokkaan, on

$$P(S_j) = \pi_{1j} + \cdots + \pi_{Ij} = \pi_{.j}.$$

Hypoteesi (13.5.11) voidaan lausua siis seuraavasti:

$$H_0\colon \pi_{ij} = \pi_{i.}\pi_{.j}\,; \qquad 1 \le i \le I, \quad 1 \le j \le J.$$

Silloin tuntemattomat parametrit ovat

$$\pi_{i.} = P(R_i), \qquad 1 \le i \le I$$

ja

$$\pi_{.j} = P(S_j), \qquad 1 \le j \le J.$$

Koska $\sum \pi_{i.} = 1$ ja $\sum \pi_{.j} = 1$, niin funktionaalisesti riippumattomien parametrien lukumäärä $q = (I-1)+(J-1)$. Siksi H_0:n testaamisessa vapausasteiden lukumäärä on

$$(k-1) - q = IJ - 1 - (I-1) - (J-1) = (I-1)(J-1).$$

$\square$

Koska H_0:n vallitessa $\pi_{ij} = \pi_{i.}\pi_{.j}$, niin logaritmoitu uskottavuusfunktio on

$$\begin{aligned}\sum_i \sum_j f_{ij} \log \pi_{ij} &= \sum_i \sum_j f_{ij}(\log \pi_{i.} + \log \pi_{.j}) \\ &= \sum_i \Big[\Big(\sum_j f_{ij}\Big) \log \pi_{i.}\Big] + \sum_j \Big[\Big(\sum_i f_{ij}\Big) \log \pi_{.j}\Big] \\ &= \sum_i f_{i.} \log \pi_{i.} + \sum_j f_{.j} \log \pi_{.j}. \qquad (13.5.12)\end{aligned}$$

Kun funktio (13.5.12) maksimoidaan rajoitteiden $\sum_i \pi_{i.} = 1$ ja $\sum_j \pi_{.j} = 1$ vallitessa, saadaan suurimman uskottavuuden estimaateiksi (H_0:n vallitessa)

$$\tilde{\pi}_{i.} = \frac{f_{i.}}{n} \qquad \text{ja} \qquad \tilde{\pi}_{.j} = \frac{f_{.j}}{n}.$$

Siksi odotettujen frekvenssien estimaatit H_0:n vallitessa ovat

$$e_{ij} = n\tilde{\pi}_{ij} = n\tilde{\pi}_{i.}\tilde{\pi}_{.j} = \frac{f_{i.}f_{.j}}{n}. \qquad (13.5.13)$$

Esimerkki 13.14 Tarkastellaan nyt taulukon 13.6 aineistoa, joka on saatu Yhdysvaltain vuoden 1991 yleisestä väestökyselystä (Agresti 1996, s. 31). Taulukossa on annettu havaitut frekvenssit ja sulkeissa odotettujen frekvens-

Taulukko 13.6. Puoluekanta sukupuolen mukaan.

	Puoluekanta			
Sukupuoli	Demokraatti	Riippumaton	Republikaani	Yhteensä
Naiset	279	73	235	577
	(261.4)	(70.7)	(244.9)	
Miehet	165	47	191	403
	(182.6)	(49.3)	(171.1)	
Yhteensä	444	120	416	980

sien estimaatit. Kaavan (13.5.13) mukaan esimerkiksi

$$e_{11} = \frac{577 \cdot 444}{980} = 261.4 \qquad \text{ja} \qquad e_{23} = \frac{403 \cdot 416}{980} = 171.1.$$

Uskottavuustestisuureen arvoksi saadaan 7.0026. Riippumattomuushypoteesin vallitessa W noudaattaa likimain χ^2-jakaumaa vapausastein $(I-1)(J-1) = (2-1)(3-1) = 2$, joten $\alpha_h \approx P(\chi_2^2 \geq 7.0026) = 0.0302$.

Pearsonin χ^2-suure

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

on tavallisimmin käytetty testisuure riippumattomuuden testaamisessa. Sekä W että Pearsonin χ^2-suure noudattavat H_0:n vallitessa asymptoottisesti samaa jakaumaa. Esimerkiksi taulukon 13.6 aineistosta $X^2 = 7.0095$ ja silloin $\alpha_h \approx P(\chi_2^2 \geq 7.0095) = 0.0301$. Tässä tapauksessa W ja Pearsonin χ^2-suure antavat melko tarkkaan saman tuloksen, koska $n = 980$ on riittävän suuri.

Sukupuolen ja puoluekannan välillä on tilastollisesti merkitsevä riippuvuus. Vertailemalla havaittuja frekvenssejä estimoituihin odotettuihin frekvensseihin saadaan käsitys tämän riippuvuuden luonteesta. Koska suurilla frekvenssien f_{ij} arvoilla myös erotus $f_{ij}-e_{ij}$ saa nollahypoteesin vallitessa itseisarvoltaan suurempia arvoja kuin pienillä frekvensseillä, ei raakaresiduaali $f_{ij}-e_{ij}$ ole riittävä poikkeamien mitta. Riippumattomuuden tarkastelussa käyttökelpoiset soluresiduaalit ovat muotoa

$$\frac{f_{ij}-e_{ij}}{\sqrt{e_{ij}(1-\tilde{\pi}_{i.})(1-\tilde{\pi}_{.j})}}.$$

□