Fazekas István

STATISZTIKA

mobiDIÁK könyvtár

Fazekas István

STATISZTIKA

mobiDIÁK könyvtár

SOROZATSZERKESZTŐ

Fazekas István

Fazekas István

STATISZTIKA

Egyetemi jegyzet
Programtervező és alkalmazott matematikusok részére
Fejlesztés alatt álló változat!

mobiDIÁK könyvtár
Debreceni Egyetem
Informatikai Intézet

Lektor

Nagy Márta
Debreceni Egyetem
Lektorálás alatt!

mobiDIÁK könyvtár
Debreceni Egyetem
Informatikai Kar
4010 Debrecen, Pf. 12
http://mobidiak.inf.unideb.hu

A mű *A mobiDIÁK önszervező mobil portál* (IKTA, OMFB-00373/2003) és a *GNU Iterátor, a legújabb generációs portál szoftver* (ITEM, 50/2003) projektek keretében készült.

Tartalomjegyzék

I. fejezet
Szórásanalízis

1. A Fisher-Cochran-tétel

A szórásanalízis alapvető tételei a Fisher-Cochran-tétel segítségével bizonyíthatók. Ebben a fejezetben a Fisher-Cochran-tételnek a nem-centrált χ^2-eloszlásokat kezelő alakját igazoljuk. Ez alapján pontosan megadható majd a szórásfelbontásban szereplő kvadratikus formák eloszlása. Lineáris algebrai ismeretekkel kezdünk, hiszen minden az ortogonális mátrixokkal történő transzformációkon fog múlni.

1.1. Kvadratikus forma rangja és szabadsági foka

1.1. definíció. Legyen $\boldsymbol{x} = (x_1, \dots, x_n)^\top$ n-dimenziós vektor, $A = (a_{ij})$ pedig $n \times n$-es (valós) szimmetrikus mátrix. Azt mondjuk, hogy a

$$Q = Q(\boldsymbol{x}) = \boldsymbol{x}^\top A \boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

kvadratikus forma rangja k, ha A rangja k. □

1.2. definíció. Legyen $\boldsymbol{b}_1, \dots, \boldsymbol{b}_m$ egy m tagú vektorrendszer $\mathbb{R}^n$-ben, $y_i = \boldsymbol{b}_i^\top \boldsymbol{x}$, $i = 1, \dots, m$, ahol $\boldsymbol{x} \in \mathbb{R}^n$. Azt mondjuk, hogy a $Q = \sum_{i=1}^{m} y_i^2 = \sum_{i=1}^{m} \left(\boldsymbol{b}_i^\top \boldsymbol{x}\right)^2$ **kvadratikus forma szabadsági foka** k, ha a $\boldsymbol{b}_1, \dots, \boldsymbol{b}_m$ vektorrendszer rangja k. □

1.1. tétel. *Legyen Q pozitív szemidefinit kvadratikus forma. Ekkor Q rangja megegyezik a szabadsági fokával.*

Bizonyítás. **1.** Legyen $Q = \boldsymbol{x}^\top A \boldsymbol{x}$, ahol A szimmetrikus, pozitív szemidefinit, $n \times n$-es mátrix. Az, hogy Q rangja k, definíció szerint azt jelenti, hogy A rangja k. De ekkor A-nak k db pozitív sajátértéke van: $\lambda_1, \dots, \lambda_k$; a többi sajátértéke 0. Az A mátrix előáll $A = U^\top \Lambda U$ alakban, ahol az U ortogonális mátrix sorai az A ortonormált sajátvektorai, a Λ diagonális mátrix főátlójában pedig az A sajátértékei állnak:

$$\Lambda = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

Ekkor $\boldsymbol{b}_i^\top = \sqrt{\lambda_i}\boldsymbol{u}_i^\top$, $i = 1, \dots, k$ választással (ahol $\boldsymbol{u}_i^\top$ az U mátrix i-edik sorvektora)

$$Q = \boldsymbol{x}^\top A \boldsymbol{x} = \boldsymbol{x}^\top U^\top \Lambda U \boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \left(\boldsymbol{u}_i^\top \boldsymbol{x}\right)^2 = \sum_{i=1}^{k} \left(\boldsymbol{b}_i^\top \boldsymbol{x}\right)^2 = \sum_{i=1}^{k} y_i^2.$$

Itt a $\boldsymbol{b}_1 = \sqrt{\lambda_1}\boldsymbol{u}_1, \dots, \boldsymbol{b}_k = \sqrt{\lambda_k}\boldsymbol{u}_k$ vektorrendszer rangja k, hiszen az $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ vektorok ortonormált rendszert alkotnak, és $\sqrt{\lambda_1} \neq 0, \dots, \sqrt{\lambda_k} \neq 0$.

2. Megfordítva, tegyük fel, hogy a $\boldsymbol{b}_1, \dots, \boldsymbol{b}_m$ vektorrendszer rangja k, és

$$Q = \sum_{i=1}^{m} (\boldsymbol{b}_i^\top \boldsymbol{x})^2 = \sum_{i=1}^{m} \boldsymbol{x}^\top \boldsymbol{b}_i \boldsymbol{b}_i^\top \boldsymbol{x} = \boldsymbol{x}^\top \left(\sum_{i=1}^{m} \boldsymbol{b}_i \boldsymbol{b}_i^\top \right) \boldsymbol{x}.$$

Be kell látnunk, hogy a $\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top$ $n \times n$-es mátrix rangja k.

Legyen a $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ vektorrendszer által kifeszített altér L. Ezen altér k-dimenziós. Belátjuk, hogy a $\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)$ (ún. diádösszeg) mátrix képtere éppen L. Mivel tetszőleges $\boldsymbol{x}$ vektorra

$$\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)\boldsymbol{x} = \sum_{i=1}^m \boldsymbol{b}_i\left(\boldsymbol{b}_i^\top\boldsymbol{x}\right) \in L,$$

így $\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)$ képtere benne van L-ben. Legyen most $\boldsymbol{v}$ a $\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)$ nullteréből való. Ekkor

$$0 = \boldsymbol{v}^\top\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)\boldsymbol{v} = \sum_{i=1}^m \left(\boldsymbol{b}_i^\top\boldsymbol{v}\right)^2$$

miatt $\boldsymbol{v} \perp \boldsymbol{b}_i$, minden i-re. Tehát $\boldsymbol{v}$ eleme L ortogonális komplementerének. Viszont mind a képtér és a nulltér dimenziójának összege, mind az altér és ortogonális komplementere dimenziójának összege n. Azaz $\left(\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^\top\right)$ képtere éppen L. □

1.1. megjegyzés. Mivel egy kvadratikus forma egyértelműen meghatározza a szimmetrikus mátrixát, így a kvadratikus forma rangja egyértelműen definiált. Viszont egy kvadratikus forma többféleképpen előállítható lineáris formák négyzetösszegeként. Tehát a szabadsági fok egyértelműségét nem közvetlenül a definícióból, hanem a szabadsági fok és a rang egyenlőségét kimondó tételből olvashatjuk ki. □

1.1. példa. A

$$Q = Q(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} = (x_1, x_2)\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = 2x_1^2 + 2x_1x_2 + x_2^2$$

kvadratikus forma rangja 2, mivel a mátrixának rangja 2.

Állítsuk elő Q-t lineáris formák négyzetösszegeként kétféle módon. Legyen először $\boldsymbol{b}_1 = (1,1)^\top$, $\boldsymbol{b}_2 = (1,0)^\top$. Nyilván

$$\left(\boldsymbol{b}_1^\top\boldsymbol{x}\right)^2 + \left(\boldsymbol{b}_2^\top\boldsymbol{x}\right)^2 = (x_1+x_2)^2 + x_1^2 = Q(\boldsymbol{x}).$$

Másrészt, $\boldsymbol{c}_1 = (\sqrt{2},\ 1/\sqrt{2})^\top$ és $\boldsymbol{c}_2 = (0,\ 1/\sqrt{2})^\top$ választással

$$\left(\boldsymbol{c}_1^\top\boldsymbol{x}\right)^2 + \left(\boldsymbol{c}_2^\top\boldsymbol{x}\right)^2 = (\sqrt{2}x_1 + x_2/\sqrt{2})^2 + (x_2/\sqrt{2})^2 = Q(\boldsymbol{x}).$$

Mind a $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, mind a $\boldsymbol{c}_1$, $\boldsymbol{c}_2$ vektorrendszer rangja 2. □

1.2. A kvadratikus forma négyzetösszeg alakja

1.2. megjegyzés. Tegyük fel, hogy a Q kvadratikus forma előáll lineáris formák négyzetösszegeként a

$$Q = \left(\boldsymbol{b}_1^\top\,\boldsymbol{y}\right)^2 + \cdots + \left(\boldsymbol{b}_s^\top\,\boldsymbol{y}\right)^2$$

és a

$$Q = \left(\boldsymbol{c}_1^\top\,\boldsymbol{y}\right)^2 + \cdots + \left(\boldsymbol{c}_r^\top\,\boldsymbol{y}\right)^2$$

alakban is. Ekkor a $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_s$ vektorrendszer által kifeszített altér megegyezik a $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_r$ vektorrendszer által kifeszítettel.

Ennek bebizonyítására tekintsük a $Q = \boldsymbol{y}^\top \left(\sum_{i=1}^s \boldsymbol{b}_i \boldsymbol{b}_i^\top\right) \boldsymbol{y}$ és a $Q = \boldsymbol{y}^\top \left(\sum_{i=1}^r \boldsymbol{c}_i \boldsymbol{c}_i^\top\right) \boldsymbol{y}$ előállításokat. Ezeknek a mátrixa megegyezik. Tehát $\sum_{i=1}^s \boldsymbol{b}_i \boldsymbol{b}_i^\top = \sum_{i=1}^r \boldsymbol{c}_i \boldsymbol{c}_i^\top$. Viszont az első mátrix képtere a $\boldsymbol{b}_1, \dots, \boldsymbol{b}_s$ vektorrendszer által kifeszített altér, míg a másodiké a $\boldsymbol{c}_1, \dots, \boldsymbol{c}_r$ vektorrendszer által kifeszített. □

1.2. példa. Állítsuk elő az $y_1^2 + y_2^2$ négyzetösszeget két kvadratikus forma összegeként:

$$y_1^2 + y_2^2 = Q_1 + Q_2 = \frac{1}{2}(y_1 + y_2)^2 + \frac{1}{2}(y_1 - y_2)^2.$$

Itt $Q_1 = (y_1/\sqrt{2} + y_2/\sqrt{2})^2$ és $Q_2 = (y_1/\sqrt{2} - y_2/\sqrt{2})^2$. A Q_1 és a Q_2 kvadratikus formák az $\boldsymbol{y} = (y_1, y_2)^\top$ vektor lineáris formáinak négyzetösszegei:

$$Q_1 = z_1^2, \quad \text{ahol} \quad z_1 = (1/\sqrt{2}, 1/\sqrt{2})\,(y_1, y_2)^\top = \boldsymbol{b}_1^\top \boldsymbol{y},$$

$$Q_2 = z_2^2, \quad \text{ahol} \quad z_2 = (1/\sqrt{2}, -1/\sqrt{2})\,(y_1, y_2)^\top = \boldsymbol{b}_2^\top \boldsymbol{y}.$$

A $\boldsymbol{z} = (z_1, z_2)^\top$ vektorra

$$\boldsymbol{z} = A\boldsymbol{y} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

tehát az új z_1, z_2 változókat a régi y_1, y_2 változókból az A ortogonális mátrixszal való transzformációval kapjuk meg. □

A fenti tény sokkal általánosabban is érvényes.

1.1. lemma. *Legyen $\boldsymbol{y} = (y_1, \dots, y_n)^\top$. Legyenek $Q_1, \dots, Q_s$ az $y_1, \dots, y_n$ változók kvadratikus formái. Tegyük fel, hogy* $\operatorname{rang}(Q_i) = n_i$, $i = 1, \dots, s$ *és*

$$\sum_{i=1}^n y_i^2 = Q_1 + \dots + Q_s.$$

Akkor és csak akkor létezik olyan A ortogonális mátrix, melyre $\boldsymbol{z} = A\boldsymbol{y}$, $\boldsymbol{z} = (z_1, \dots, z_n)^\top$ jelöléssel

$$Q_1 = z_1^2 + \dots + z_{n_1}^2, \quad Q_2 = z_{n_1+1}^2 + \dots + z_{n_1+n_2}^2, \ \dots,$$

$$Q_s = z_{n_1+\dots+n_{s-1}+1}^2 + \dots + z_{n_1+\dots+n_s}^2$$

teljesül, ha $n_1 + n_2 + \dots + n_s = n$.

Bizonyítás. **1.** Tegyük fel, hogy létezik a fenti A mátrix. Ekkor $A^\top A = I$ miatt

$$\sum_{i=1}^n z_i^2 = \boldsymbol{z}^\top \boldsymbol{z} = \boldsymbol{y}^\top A^\top A \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{y} = \sum_{i=1}^n y_i^2 = \sum_{j=1}^s Q_j = \sum_{i=1}^{n_1+\dots+n_s} z_i^2$$

minden $\boldsymbol{z} \in \mathbb{R}^n$-re. Ezért $n = n_1 + \dots + n_s$.

2. Tegyük fel, hogy $n_1 + \dots + n_s = n$. Mivel Q_1 n_1 rangú kvadratikus forma, így „teljes négyzetté alakításában” n_1 db lineáris forma szerepel:

$$Q_1 = \sum\nolimits_{i=1}^{n_1} \delta_i z_i^2,$$

ahol $\delta_i = \pm 1$, $z_i = \boldsymbol{b}_i^\top \boldsymbol{y}$, $i = 1, \dots, n_1$. Hasonló igaz $Q_2, \dots, Q_s$-re is. Mivel $n_1 + \dots + n_s = n$, így a Q_i kvadratikus formák előállításában szereplő $\boldsymbol{b}_j^\top$ n-dimenziós sorvektorokat mind egymás alá írva, egy $n \times n$-es mátrixot kapunk:

$$A = \begin{pmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_{n_1}^\top \\ \vdots \\ \boldsymbol{b}_n^\top \end{pmatrix}.$$

Jelölje most D a δ_i diagonális elemekből álló diagonális mátrixot (δ_i-ket ugyanolyan sorrendben írjuk le, mint a megfelelő $\boldsymbol{b}_i^\top$ vektorokat az A mátrixban). Ekkor

$$\boldsymbol{y}^\top A^\top D A \boldsymbol{y} = \boldsymbol{z}^\top D \boldsymbol{z} = \sum_{i=1}^{n} \delta_i z_i^2 = \sum_{j=1}^{s} Q_j = \sum_{i=1}^{n} y_i^2 = \boldsymbol{y}^\top \boldsymbol{y}$$

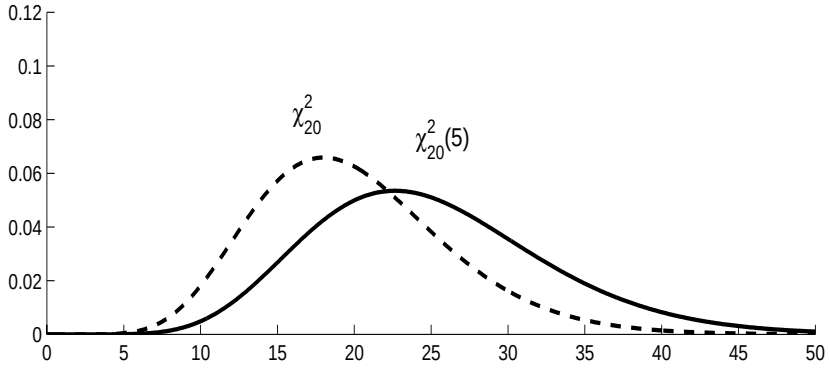
minden $\boldsymbol{y} \in \mathbb{R}^n$-re.

Mivel egy kvadratikus forma egyértelműen meghatározza a szimmetrikus mátrixát, ezért $A^\top D A = I$. Ebből adódik, hogy A invertálható, és $D = (A^\top)^{-1} A^{-1} = (A^{-1})^\top A^{-1}$. Tehát D egy pozitív szemidefinit mátrixszal egyenlő, ami alapján a főátlójában szereplő δ_i-k egyike sem lehet -1, azaz $D = I$. De ekkor $A^\top A = I$, azaz A ortogonális. Továbbá

$$Q_1 = \sum\nolimits_{i=1}^{n_1} z_i^2,$$

és hasonló áll fenn a többi Q_j-re is. □

1.3. A nem-centrált khi-négyzet eloszlás



ÁBRA 1.1. χ^2_{20} és $\chi^2_{20}(5)$ sűrűségfüggvénye

1.3. definíció. Legyenek $\xi_1, \dots, \xi_n$ független, normális eloszlású valószínűségi változók: $\xi_i \sim \mathcal{N}(a_i, 1)$, $i = 1, \dots, n$. Ekkor a

$$\zeta_n = \xi_1^2 + \dots + \xi_n^2 \tag{1.1}$$

valószínűségi változót n szabadsági fokú, $\lambda = \sum_{i=1}^n a_i^2$ nem-centralitási paraméterű **nem-centrált khi-négyzet eloszlásúnak** nevezzük. Jelölése $\chi_n^2(\lambda)$. □

A nem-centrált khi-négyzet eloszlás abban a speciális esetben, amikor a kiinduló ξ_i valószínűségi változók 0 várható értékűek, éppen a korábban megismert (centrált) khi-négyzet eloszlás. Azaz $\chi_n^2(0) \equiv \chi_n^2$.

A χ_{20}^2 és $\chi_{20}^2(5)$ sűrűségfüggvénye a 1.1. ábrán látható.

A következő állítás azt mutatja, hogy nem kell a ξ_i-k a_i várható értékeit külön-külön ismerni, az eloszlás csak a $\lambda = \sum_{i=1}^n a_i^2$ nem-centralitási paramétertől függ.

1.1. állítás. *Legyenek $\tau_1, \dots, \tau_n$ független, normális eloszlású valószínűségi változók: $\tau_1 \sim \mathcal{N}(\sqrt{\lambda}, 1)$, $\tau_2 \sim \mathcal{N}(0,1), \dots, \tau_n \sim \mathcal{N}(0,1)$, ahol $\lambda \geq 0$. Ekkor*

$$\tau_1^2 + \dots + \tau_n^2 \tag{1.2}$$

eloszlása megegyezik az (1.1)-ben adott ζ_n eloszlásával, bármilyenek is a $\lambda = \sum_{i=1}^n a_i^2$ feltételt kielégítő $a_1, \dots, a_n$ számok.

Bizonyítás. Legyen a $B = (b_{ij})$ $n \times n$-es ortogonális mátrix első sora:

$$b_{11} = a_1/\sqrt{\lambda}, \ \dots \ , b_{1n} = a_n/\sqrt{\lambda}\,.$$

Ilyen B mátrix létezik, hisz az első sorra előírt vektor egységnyi hosszúságú, ez kiegészíthető ortonormált bázissá, és B sorai a bázisvektorok lesznek.

Legyen $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top$, ahol $\xi_1, \dots, \xi_n$ a 1.3. definícióbeliek. Legyen $\boldsymbol{\tau} = B\boldsymbol{\xi}$. Mivel $\boldsymbol{\xi}$ eloszlása $\mathcal{N}_n(\boldsymbol{m}, I)$, ahol $\boldsymbol{m} = (a_1, \dots, a_n)^\top$, így $\boldsymbol{\tau} \sim \mathcal{N}_n(B\boldsymbol{m}, I)$. Ezért $\boldsymbol{\tau}$ koordinátái, $\tau_1, \dots, \tau_n$, együttesen normális eloszlású és korrelálatlan valószínűségi változók. Így függetlenek is. A várható értékeik pedig a $B\boldsymbol{m}$ vektor koordinátái. Azaz $\mathbb{E}\tau_i = \boldsymbol{b}_i^\top \boldsymbol{m}$, $i = 1, \dots, n$, ahol $\boldsymbol{b}_i^\top$ a B mátrix i-edik sora. De $\boldsymbol{b}_1 = \lambda^{-\frac{1}{2}}\boldsymbol{m}$, továbbá $\boldsymbol{b}_2, \dots, \boldsymbol{b}_n$ erre merőleges. Ezért

$$\mathbb{E}\tau_1 = \lambda^{-\frac{1}{2}}\boldsymbol{m}^\top\boldsymbol{m} = \lambda^{-\frac{1}{2}}\sum_{i=1}^n a_i^2 = \lambda^{-\frac{1}{2}}\lambda = \sqrt{\lambda}\,.$$

$$\mathbb{E}\tau_i = \boldsymbol{b}_i^\top \boldsymbol{b}_1 \sqrt{\lambda} = 0, \quad i = 2, \dots, n.$$

Tehát $\tau_1, \dots, \tau_n$ kielégíti a tétel feltételeit. Másrészt

$$\sum\nolimits_{i=1}^n \xi_i^2 = \boldsymbol{\xi}^\top\boldsymbol{\xi} = \boldsymbol{\xi}^\top B^\top B\boldsymbol{\xi} = \boldsymbol{\tau}^\top\boldsymbol{\tau} = \sum\nolimits_{i=1}^n \tau_i^2.$$

Tehát az (1.2) előállítás teljesül. □

1.2. állítás. *A nem-centrált $\chi_n^2(\lambda)$ khi-négyzet eloszlás karakterisztikus függvénye:*

$$\varphi(t) = (1-2it)^{-n/2}\exp[it\lambda(1-2it)^{-1}];$$

várható értéke:

$$\mathbb{E}\zeta_n = n + \lambda;$$

szórásnégyzete:

$$\mathbb{D}^2\zeta_n = 2(n+2\lambda);$$

ferdesége:

$$\beta_1(\zeta_n) = \frac{\sqrt{8}(n+3\lambda)}{(n+2\lambda)^{3/2}};$$

lapultsága:

$$\beta_2(\zeta_n) = \frac{12(n+4\lambda)}{(n+2\lambda)^2}.$$

BIZONYÍTÁS. Legyen $\xi \sim \mathcal{N}(m,1)$. Ekkor ξ^2 karakterisztikus függvénye:

$$\varphi_{\xi^2}(t) = \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2}}\, dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(x\sqrt{1-2it} - \frac{m}{\sqrt{1-2it}}\right)^2\right\} dx \exp\left\{\frac{itm^2}{1-2it}\right\} =$$

$$= \frac{1}{\sqrt{1-2it}} \exp\left\{\frac{itm^2}{1-2it}\right\},$$

ahol az utolsó lépésben az

$$y = x\sqrt{1-2it} - \frac{m}{\sqrt{1-2it}}$$

helyettesítést végeztük el. Az így kapott karakterisztikus függvényekből szorzással adódik $\chi^2_n(\lambda)$ karakterisztikus függvénye.

Ha $\xi \sim \mathcal{N}(m,\sigma^2)$, akkor

$$\mathbb{E}\xi^2 = \sigma^2 + m^2, \qquad \mathbb{D}^2\xi^2 = 2\sigma^4 + 4\sigma^2 m^2$$

(ez utóbbit a standard normálisra visszavezetve kaphatjuk). Ezután alkalmazzuk $\chi^2_n(\lambda)$ definícióját! □

1.2. tétel. *(χ^2 addíciós tétel.) Legyenek η_n és η_m független χ^2-eloszlású valószínűségi változók n, ill. m szabadsági fokkal, valamint λ_1, ill. λ_2 nem-centralitási paraméterrel. Ekkor az $\eta_n + \eta_m$ változó $n+m$ szabadsági fokú, $\lambda_1+\lambda_2$ nem-centralitási paraméterű χ^2-eloszlású.*

A tétel szavakban kifejezve: független χ^2-ek összege χ^2, a szabadsági fokok és a nem-centralitási paraméterek pedig összeadódnak.

BIZONYÍTÁS. Legyenek $\xi_1, \ldots, \xi_{n+m}$ független normális eloszlásúak: $\xi_i \sim \mathcal{N}(a_i, 1)$, $i = 1, \ldots, n+m$. η_n-et $\xi_1^2 + \cdots + \xi_n^2$-ként, η_m-et pedig $\xi_{n+1}^2 + \cdots + \xi_{n+m}^2$-ként reprezentálva,

$$\eta_n + \eta_m = \xi_1^2 + \cdots + \xi_{n+m}^2$$

adódik, ez pedig $n+m$ szabadsági fokú, $\lambda_1+\lambda_2$ nem-centralitási paraméterű khi-négyzet eloszlású. □

1.4. definíció. Legyenek ξ és η független χ^2-eloszlású valószínűségi változók: $\xi_i \sim \chi^2_n(\lambda)$, $\eta_i \sim \chi^2_m$. Ekkor a

$$\frac{\xi}{n} \Big/ \frac{\eta}{m} \tag{1.3}$$

valószínűségi változót n és m szabadsági fokú, λ nem-centralitási paraméterű **nem-centrált F-eloszlásúnak** nevezzük. □

A nem-centrált F-eloszlás abban a speciális esetben, amikor a kiinduló ξ valószínűségi változó is centrált, éppen a korábban megismert (centrált) F-eloszlás.

1.4. A Fisher-Cochran-tétel

1.3. tétel. Fisher-Cochran. *Legyenek $\xi_1, \dots, \xi_n$ független valószínűségi változók, $\xi_i \sim \mathcal{N}(a_i, 1)$, $i = 1, \dots, n$. Legyenek $Q_1, \dots, Q_s$ a $\xi_1, \dots, \xi_n$ kvadratikus formái,* $\operatorname{rang}(Q_i) = n_i$, $i = 1, \dots, s$. *Tegyük fel, hogy*

$$\xi_1^2 + \dots + \xi_n^2 = Q_1 + \dots + Q_s.$$

Ekkor $Q_1, \dots, Q_s$ akkor és csak akkor függetlenek és $\chi^2_{n_1}, \dots, \chi^2_{n_s}$ eloszlásúak, ha $n_1 + \dots + n_s = n$.

Ebben az esetben Q_i nem-centralitási paramétere: $Q_i(\boldsymbol{a})$, ahol $\boldsymbol{a} = (a_1, \dots, a_n)^\top$.

Bizonyítás. **1.** Legyenek $Q_1, \dots, Q_s$ függetlenek, Q_i eloszlása $\chi^2_{n_i}$, $i = 1, \dots, s$. Ekkor a χ^2-addíciós tétel alapján $Q_1 + \dots + Q_s$ eloszlása $\chi^2_{n_1+\dots+n_s}$. De $Q_1 + \dots + Q_s = \xi_1^2 + \dots + \xi_n^2$, aminek az eloszlása definíció szerint χ^2_n. Tehát $n_1 + \dots + n_s = n$. Megjegyezzük, hogy itt nem volt külön szükség a $\operatorname{rang}(Q_i) = n_i$ feltételre.

2. Legyen most $n_1 + \dots + n_s = n$. Az 1.1. lemma miatt létezik olyan A ortogonális mátrix, hogy az $\boldsymbol{\eta} = A\boldsymbol{\xi}$-re (ahol $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^\top$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top$)

$$Q_1 = \eta_1^2 + \dots + \eta_{n_1}^2, \quad Q_2 = \eta_{n_1+1}^2 + \dots + \eta_{n_1+n_2}^2, \dots,$$

$$Q_s = \eta_{n_1+\dots+n_{s-1}+1}^2 + \dots + \eta_{n_1+\dots+n_s}^2.$$

Viszont $\boldsymbol{\xi} \sim \mathcal{N}_n(\boldsymbol{a}, I)$, ezért $\boldsymbol{\eta} = A\boldsymbol{\xi} \sim \mathcal{N}_n(A\boldsymbol{a}, AA^\top) = \mathcal{N}_n(A\boldsymbol{a}, I)$, hiszen A ortogonális mátrix.

Tehát $\boldsymbol{\eta}$ koordinátái független, 1 szórású normálisak. Q_i pedig n_i db ilyennek a négyzetösszege, tehát $\chi^2_{n_i}$ eloszlású. Mivel különböző Q_i-k előállításában azonos η_j-k nem vesznek részt, így Q_i-k függetlenek.

3. A nem-centralitási paraméter. A lemmában adott előállítás $Q_1 = z_1^2 + \dots + z_{n_1}^2$; részletesen kiírva:

$$Q_1(\boldsymbol{y}) = (\boldsymbol{b}_1^\top \boldsymbol{y})^2 + \dots + (\boldsymbol{b}_{n_1}^\top \boldsymbol{y})^2, \quad \boldsymbol{y} \in \mathbb{R}^n,$$

ahol $\boldsymbol{b}_1^\top, \dots, \boldsymbol{b}_n^\top$ az A mátrix sorvektorai. Ezt a $\boldsymbol{\xi}$ várható érték vektora, azaz $\boldsymbol{y} = \boldsymbol{a}$ helyén véve:

$$Q_1(\boldsymbol{a}) = (\boldsymbol{b}_1^\top \boldsymbol{a})^2 + \dots + (\boldsymbol{b}_{n_1}^\top \boldsymbol{a})^2 = (\mathbb{E}\eta_1)^2 + \dots + (\mathbb{E}\eta_{n_1})^2.$$

Ez utóbbi pedig, a $Q_1 = \eta_1^2 + \dots + \eta_{n_1}^2$ alapján, éppen Q_1 nem-centralitási paramétere. Hasonló áll $Q_2, \dots, Q_s$-re. □

1.3. megjegyzés. A fenti bizonyításból kiderült, hogy a Fisher-Cochran-tétel feltételei esetén a $\xi_1^2 + \dots + \xi_n^2$ teljes négyzetösszeg előáll a $\xi_1, \dots, \xi_n$ lineáris formái négyzetösszegeként:

$$\xi_1^2 + \dots + \xi_n^2 = (\boldsymbol{b}_1^\top \boldsymbol{\xi})^2 + \dots + (\boldsymbol{b}_n^\top \boldsymbol{\xi})^2,$$

ahol az $\eta_i = \boldsymbol{b}_i^\top \boldsymbol{\xi}$, $i = 1, \dots, n$, lineáris formák függetlenek, és $\eta_i \sim \mathcal{N}(\boldsymbol{b}_i^\top \boldsymbol{a}, 1)$. Sőt, Q_j ezek közül az első n_1 db, Q_2 a következő n_2 db, … négyzetösszege.

Ha Fisher-Cochran-tétel feltételei mellett létezik a Q_i kvadratikus formáknak valamilyen lineáris formák négyzetösszegeként történő előállítása, akkor a különböző Q_i-k előállításában szereplő formák függetlenek.

Ezt bizonyítsuk arra az esetre, ha Q_1 előáll a $Q_1 = (\boldsymbol{b}_1^\top \boldsymbol{\xi})^2 + \cdots + (\boldsymbol{b}_{n_1}^\top \boldsymbol{\xi})^2$ mellett $Q_1 = (\boldsymbol{c}_1^\top \boldsymbol{\xi})^2 + \cdots + (\boldsymbol{c}_m^\top \boldsymbol{\xi})^2$ alakban is. Egy előző megjegyzés szerint a $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$ által kifeszített altér megegyezik a $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n_1}$ által kifeszítettel. De az utóbbi ortogonális a $\boldsymbol{b}_{n_1+1}, \ldots, \boldsymbol{b}_n$ által kifeszített altérre. Ezért $\boldsymbol{c}_i^\top \boldsymbol{b}_j = 0$, amiből

$$\operatorname{cov}(\boldsymbol{c}_i^\top \boldsymbol{\xi}, \boldsymbol{b}_j^\top \boldsymbol{\xi}) = \boldsymbol{c}_i^\top I \boldsymbol{b}_j = \boldsymbol{c}_i^\top \boldsymbol{b}_j = 0,$$

ha $1 \le i \le m$, $n_1 + 1 \le j \le n$. A $(\boldsymbol{c}_1^\top \boldsymbol{\xi}, \ldots, \boldsymbol{c}_m^\top \boldsymbol{\xi})^\top$ és a $(\boldsymbol{b}_{n_1+1}^\top \boldsymbol{\xi}, \ldots, \boldsymbol{b}_n^\top \boldsymbol{\xi})^\top$ vektorok együttesen normális eloszlásúak, korrelálatlanok, tehát függetlenek. Ezért pl. $\boldsymbol{c}_1^\top \boldsymbol{\xi}$ és Q_2 független. □

2. Az egyszeres osztályozás

Ebben a szakaszban a Fisher-Cochran-tételt alkalmazzuk az egyszeres osztályozásra.

2.1. Az egyszeres osztályozás feladata

A szórásanalízis (ANOVA=ANalysis Of VAriance) alapkérdése: több minta esetén a várható értékek egyenlőek-e. Alapvető feltétel: minták egymástól is függetlenek, normális eloszlásból származnak, és a szórásaik egyenlőek! Tehát a minták között csak a várható értékeikben lehet eltérés.

A legegyszerűbb szórásanalízisbeli modell az egyszeres osztályozás (one-way classification, one-way layout). Itt egyetlen **tényező szintjei**t kell összehasonlítani. Mivel a megfigyelések eredményeit tényezőnként egy-egy oszlopban szokták elhelyezni, a tényezők szintjeinek hatását oszlophatásnak nevezzük. Példaként tekintsünk egy fogyókúrás kísérletet.

diéta tényező		
1. diéta	2. diéta	3. diéta
9.40	22.84	17.35
9.48	15.32	16.36
7.56	11.04	15.88
11.52	17.92	14.28
11.56	19.68	18.60
12.12	26.20	19.32
11.36		14.20
4.60		17.52
14.48		

2.1. példa. Három különböző diéta hatását mérték 9, 6, ill. 8 kísérleti alanyon. Itt az egyetlen tényező a diéta, annak 3 szintje van. A diéta hatására a súlyveszteségeket a fenti táblázat mutatja. Vizsgáljuk meg azt a nullhipotézist, hogy a diéták által okozott súlyveszteségek várható értékei egyenlőek! □

A megfigyelések: Y_{ij}, $j = 1, \dots, n_i$, $i = 1, \dots, p$. Y_{ij} az i-edik szinten végzett j-edik megfigyelés. Az egyes szinteken nem feltétlen kell azonos számú mérést végezni.

Feltesszük hogy,

$$Y_{ij} \sim \mathcal{N}(\mu + \alpha_i, \sigma^2) \quad \text{és} \quad Y_{ij}\text{-k} \quad \text{függetlenek.}$$

Elérhető, hogy $\sum_{i=1}^{p} n_i \alpha_i = 0$ legyen. Vezessük be az $n = n_1 + \dots + n_p$ jelölést.

Vizsgáljuk a

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

nullhipotézis teljesülését! H_0 azt jelenti, hogy az egyes szinteknek nincs hatása.

A Steiner-formula alapján az n db Y_{ij} négyzetösszege előáll

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}^2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 + n\overline{Y}_{\cdot\cdot}^2$$

alakban, ahol $\overline{Y}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}$ a **teljes átlag**. A fenti felbontásban szereplő első négyzetösszeg jelölése Q, elnevezése **teljes négyzetösszeg**. Q előáll

$$Q = Q_1 + Q_2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i} \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{i=1}^{p}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$$

alakban, ahol $\overline{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ az i-edik **szint átlaga**.

Q_1 méri a **szintek közötti szóródás**t, Q_2 pedig a **szinteken belüli szóródás**t (azaz a véletlen hibát). H_0-at akkor vetjük el, ha Q_1 túlságosan nagy Q_2-höz képest.

2.2. A Fisher-Cochran-tétel alkalmazása az egyszeres osztályozásra

A próba konstruálására használjuk a Fisher-Cochran-tételt. Tekintsük a teljes felbontást:

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}^2 = Q_1 + Q_2 + n\overline{Y}_{\cdot\cdot}^2 .$$

2.1. állítás. *Q_1 szabadsági foka $p-1$, Q_2 szabadsági foka $n-p$, $n\overline{Y}_{\cdot\cdot}^2$ szabadsági foka* 1.

BIZONYÍTÁS. A kvadratikus formák itt lineáris formák négyzetösszegeiként vannak előállítva. Rendezzük a lineáris formák együtthatóit

$$Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots, Y_{p1}, \ldots, Y_{pn_p}$$

szerint.

Ekkor a Q_1-hez tartozó együtthatók mátrixa (valójában az azonos sorok közül $1-1$-et megtartva):

$$\begin{pmatrix} \frac{1}{n_1}-\frac{1}{n} & \ldots & \frac{1}{n_1}-\frac{1}{n} & -\frac{1}{n} & \ldots & -\frac{1}{n} & \ldots & -\frac{1}{n} & \ldots & -\frac{1}{n} \\ -\frac{1}{n} & \ldots & -\frac{1}{n} & \frac{1}{n_2}-\frac{1}{n} & \ldots & \frac{1}{n_2}-\frac{1}{n} & \ldots & -\frac{1}{n} & \ldots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & & & & & & & \\ -\frac{1}{n} & \ldots & -\frac{1}{n} & -\frac{1}{n} & \ldots & -\frac{1}{n} & \ldots & \frac{1}{n_p}-\frac{1}{n} & \ldots & \frac{1}{n_p}-\frac{1}{n} \end{pmatrix}.$$

Ezen p sor lineárisan függő, hisz $n_1, \ldots, n_p$ együtthatókkal belőlük 0 kombinálható. De az elsőt kivonva az összes többiből, $p-1$ lineárisan független sort kapunk. Azaz Q_1 szabadsági foka $p-1$.

A Q_2-ben szereplő lineáris formák együtthatói olyan

$$D = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_p \end{pmatrix}$$

mátrixot alkotnak, melyben a diagonálison fekvő $D_1, \dots, D_p$ blokkokon kívüli elemek 0-k. A D_i blokk:

$$D_i = \begin{pmatrix} 1-\frac{1}{n_i} & -\frac{1}{n_i} & \dots & -\frac{1}{n_i} \\ -\frac{1}{n_i} & 1-\frac{1}{n_i} & \dots & -\frac{1}{n_i} \\ \vdots & \vdots & & \\ -\frac{1}{n_i} & -\frac{1}{n_i} & \dots & 1-\frac{1}{n_i} \end{pmatrix}.$$

Ez az $n_i \times n_i$ méretű mátrix $n_i - 1$ rangú. Ugyanis sorainak összege 0, de az első sort az összes többiből kivonva, $n_i - 1$ db lineárisan független sort kapunk. Ezek alapján Q_2 szabadsági foka $(n_1 - 1) + \dots + (n_p - 1) = n - p$.

$n\overline{Y}_{\cdot\cdot}^2$ szabadsági foka nyilván 1. □

2.2. állítás. *Q_1 és Q_2 függetlenek. $\frac{1}{\sigma^2}Q_1 \sim \chi^2_{p-1}(\lambda_1)$, $\frac{1}{\sigma^2}Q_2 \sim \chi^2_{n-p}(0)$, ahol $\lambda_1 = \frac{1}{\sigma^2}\sum_{i=1}^{p} n_i\alpha_i^2$.*

$\frac{1}{\sigma^2}Q_1$ akkor és csak akkor centrált χ^2-eloszlású, ha a

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

nullhipotézis teljesül.

Bizonyítás. Az előző állítás miatt a Fisher-Cochran-tétel feltétele teljesül, hiszen a rangok összege $(p-1) + (n-p) + 1 = n$. Így a kvadratikus formák függetlenek és χ^2-eloszlásúak. A nem-centralitási paramétereket úgy kapjuk, hogy a várható értékeket beírjuk a kvadratikus formákba. Mivel $\mathbb{E}(Y_{ij} - \overline{Y}_{i\cdot})=0$, így Q_2 nem-centralitási paramétere 0. Viszont $\mathbb{E}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}) = \alpha_i$, ezért $\frac{1}{\sigma^2}Q_1$ nem-centralitási paramétere: $\lambda_1 = \frac{1}{\sigma^2}\sum_{i=1}^{p} n_i\alpha_i^2$. □

2.3. A szórásfelbontó táblázat

A 2.2. állítás alapján a Q_2 véletlen hiba (skalárszorosa) mindig centrált χ^2-eloszlású, Q_1 (skalárszorosa) viszont pontosan akkor, ha H_0 teljesül.

2.3. állítás. *Az*

$$F = \frac{Q_1}{p-1} \Big/ \frac{Q_2}{n-p}$$

statisztika pontosan akkor $p-1$ és $n-p$ szabadsági fokú centrált F-eloszlású, ha a

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

nullhipotézis teljesül.

Bizonyítás. Az 2.2. állítás szerint F két független, a szabadsági fokával elosztott, (H_0 esetén centrált) χ^2-eloszlású valószínűségi változó hányadosa. Ezért az F-eloszlás definíciója alapján ennek eloszlása (H_0 esetén centrált) F-eloszlás $p-1$ és $n-p$ szabadsági fokokkal. □

Az eddigiek alapján az alábbi szórásfelbontó táblázatot adhatjuk meg az egytényezős osztályozásra.

A szóródás forrásai	Szabadsági fokok	Négyzet-összegek	Négyzet-átlagok	F-hányados
Részsokaságok (szintek) közötti eltérések	$p-1$	Q_1	$\frac{Q_1}{p-1}$	$F=\frac{n-p}{p-1}\cdot\frac{Q_1}{Q_2}$
Szinteken belüli eltérések (véletlen hibák)	$n-p$	Q_2	$\frac{Q_2}{n-p}$	
Teljes	$n-1$	Q_1+Q_2		

Szórásfelbontó táblázat az egytényezős osztályozásra

H_0-at $1-\alpha$ szinten elvetjük, ha a kapott F-statisztika értéke nagyobb, mint $F_{[p-1,n-p;\alpha]}$, azaz a megfelelő szabadsági fokú centrált F-eloszlás táblázatából kikeresett (felső) kritikus érték.

2.2. példa. (A 2.1. példa folytatása.)
A (számítógépes) eredményt a szórásfelbontó tábla tartalmazza:

ANOVA Table				
Source	df	SS	MS	F
Columns	2	313.9	156.90	13.31
Error	20	235.8	11.79	
Total	22	549.7		

Az elnevezések magyarázata. Source = a szóródás forrása; Columns = oszlophatás (szintek közötti eltérések); Error = véletlen hiba; Total = teljes négyzetösszeg; df (degree of freedom) = szabadsági fok; SS (Sum of Squares) = négyzetösszeg; MS (Mean Square) = tapasztalati szórásnégyzet (négyzet átlag), F = F-statisztika.

Annak kérdéséről, hogy a diéta három szintjének van-e hatása, az F alatti mennyiség alapján döntünk. Amennyiben H_0 : *a tényező szintjeinek nincs hatása* nullhipotézis teljesül, az F alatti statisztika centrált F-eloszlású (jelenleg (2, 20) szabadsági fokkal). Ez alapján határozható meg a próba pontos terjedelme: p. Példánkban $p = 0.00021$ érték adódott, azaz minden használatos szinten elvetjük a diéták egyforma hatását.

A hagyományos (táblázatos) kiértékelés ugyanerre a következtetésre vezet. F értékét összehasonlítva a $(2\,,20)$ szabadsági fokú centrált F-eloszlás $F_{[2,20;0.05]} = 3.49$ kritikus értékével, azt kapjuk, hogy a H_0 nullhipotézist 95%-os szinten el kell vetni. Ez azt jelenti, hogy a diéta tényező különböző szintjeinek van hatásuk a súlycsökkenésre.

Megjegyezzük, hogy az eljárást formálisan végrehajtottuk, azonban az alapfeltevések nem teljesülnek. Példánkban sem a szórások nem egyenlőek, sem a normalitás nem igaz (ez utóbbi grafikus eljárások, azaz hisztogram és Gauss-papír alapján adódott). Transzformációkkal (logaritmus, illetve törtkitevős hatvány vétele) részleges javulást sikerült elérni, a transzformáció elvégzését az olvasóra bízzuk. □

2.4. Konfidencia intervallum a várható értékre

Amennyiben a

$$H_0 \,:\, \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$$

nullhipotézist elvetjük, úgy konfidencia intervallumot kell szerkesztenünk a (különböző) várható értékekre. Megmutatjuk, hogy az egyszeres osztályozás esetén hogyan lehet konfidencia intervallumot szerkeszteni a várható értékekre, ill. azok különbségeire.

A teljes négyzetösszeg felbontásában szereplő lineáris formák függetlenek. Így

$$\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot},\ \overline{Y}_{j\cdot} - \overline{Y}_{\cdot\cdot}\right), \quad Q_2, \quad \overline{Y}_{\cdot\cdot}$$

független (kettő-, illetve egy-egydimenziós) valószínűségi változók ($i \neq j$).

Emiatt $\overline{Y}_{i\cdot}$ és Q_2 függetlenek. Ezek eloszlása:

$$\overline{Y}_{i\cdot} \sim \mathcal{N}\left(\mu + \alpha_i, \frac{\sigma^2}{n_i}\right), \quad \frac{1}{\sigma^2} Q_2 \sim \chi^2_{n-p}(0)\,.$$

Így

$$\sqrt{n_i}\left(\overline{Y}_{i\cdot} - (\mu + \alpha_i)\right) / \sqrt{Q_2/(n-p)} \sim t_{n-p}\,.$$

Ez alapján (ha H_0-at elvetjük) lehet konfidencia intervallumot szerkeszteni $(\mu + \alpha_i)$-re.

Másrészt, a fentiek alapján,

$$\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot} \sim \mathcal{N}\left(\alpha_i, \left(\frac{1}{n_i} - \frac{1}{n}\right)\sigma^2\right),$$

ahol $\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$ szórása a lineáris forma együtthatói alapján adódott. Tehát ebből a lineáris formából és a Q_2 kvadratikus formából lehet α_i-re konfidencia intervallumot szerkeszteni a t_{n-p}-eloszlás alapján.

Most szerkesszünk konfidencia intervallumot a várható értékek különbségeire.

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \ \text{ és } \ Q_2$$

függetlenek. Továbbá

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \sim \mathcal{N}\left(\alpha_i - \alpha_j, \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)\right).$$

Ebből és Q_2-ből $(\alpha_i - \alpha_j)$-re lehet konfidencia intervallumot szerkeszteni, hiszen

$$\sqrt{\frac{n_i n_j}{n_i + n_j}}\left(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} - (\alpha_i - \alpha_j)\right) / \sqrt{Q_2/(n-p)} \sim t_{n-p}\,.$$

3. Kétszeres osztályozás interakció nélkül

3.1. A kétszeres osztályozás feladata

3.1. példa. Egy üzemben 5 gépen, és 3-féle anyagból gyártják ugyanazt a terméket. Jelölje Y a termék egy jellemzőjét. Azonosnak tekinthető-e Y a különböző gépeken és különböző anyagokból készített termékek esetén? □

A kétszeres osztályozás azt jelenti, hogy két tényező befolyásolja a kísérlet eredményét. Ezeket jelölje A és B. Tegyük fel, hogy az A tényezőnek I, a B tényezőnek pedig J különböző szintje van. Fenti példánkban A a gép, ennek 5 szintje az 5 gép maga. B pedig az anyag, ezen tényező 3 szintje a 3-féle anyag.

A két tényezőnek minden szintkombinációjára végzünk megfigyeléseket, a mérési eredményeket mátrixba rendezzük. Az egyik tényező szintjei a mátrix sorait, a másik tényező szintjei az oszlopait jelöli. Ezért az egyik tényezőt **sorhatás**nak, a másikat **oszlophatás**nak nevezzük. (Egy szántóföldi kísérletnél a tényleges parcellák is így helyezkedhetnek el.)

A két tényező szintkombinációinak száma, azaz a cellák száma IJ. Az elsőként tárgyalandó modellünkben feltételezzük, hogy a tényezőknek együttes hatása (interakciója) nincs. Ekkor elegendő cellánként egy megfigyelést végezni. Legyen az (i,j)-edik cellában végzett kísérlet eredménye Y_{ij}, $(i = 1,2,\dots,I;\ j = 1,2,\dots,J)$. Tegyük fel, hogy

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1,2,\dots,I, \quad j = 1,2,\dots,J\,, \tag{3.1}$$

ahol α_i az i-edik sor hatása, β_j pedig a j-edik oszlop hatása; továbbá az ε_{ij}, $i = 1,2,\dots,I$, $j = 1,2,\dots,J$, változók függetlenek, normális eloszlásúak:

$$\varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2)\,.$$

Fontos (és a tényleges adatok esetén ellenőrizendő), hogy a megfigyelések független, azonos szórású normálisak. Az egyes tényezők szintjei csak a megfigyelések várható értékét befolyásolják.

Elérhető, hogy a paraméterek kielégítsék a következő feltételeket:

$$\sum\nolimits_{i=1}^{I} \alpha_i = 0, \qquad \sum\nolimits_{j=1}^{J} \beta_j = 0. \tag{3.2}$$

A vizsgálandó hipotézisek:

$$H_A : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0\,,$$
$$H_B : \beta_1 = \beta_2 = \cdots = \beta_J = 0\,,$$

azaz nincs A-hatás, ill. nincs B-hatás.

Vezessük be a szokásos jelöléseket:

$$\overline{Y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}, \qquad i = 1, 2, \ldots, I;$$

$$\overline{Y}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} Y_{ij}, \qquad j = 1, 2, \ldots, J;$$

$$\overline{Y}_{\cdot\cdot} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij},$$

$$Q_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2, \qquad Q_2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2,$$

$$Q_3 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i\cdot} - \overline{Y}_{\cdot j} + \overline{Y}_{\cdot\cdot})^2, \qquad Q = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2.$$

Itt Q a teljes négyzetösszeg, Q_1 az A tényező szintjei közötti eltérések négyzetösszege, Q_2 a B tényező szintjei közötti eltérések négyzetösszege, végül Q_3 a hibatag. Elemi számolással megmutatható, hogy Q felbontható Q_1, Q_2 és Q_3 összegére.

3.2. A Fisher-Cochran-tétel alkalmazása a kétszeres osztályozásra

A Fisher-Cochran-tétel alkalmazása szempontjából praktikusabb az alábbi felbontásból kiindulni.

$$\sum\nolimits_{i=1}^{I}\sum\nolimits_{j=1}^{J} Y_{ij}^2 = Q_1 + Q_2 + Q_3 + IJ\,\overline{Y}_{\cdot\cdot}^{\,2}\,.$$

3.1. állítás. *Q_1 szabadsági foka $I-1$, Q_2 szabadsági foka $J-1$, Q_3 szabadsági foka $(I-1)(J-1)$, $IJ\,\overline{Y}_{\cdot\cdot}^{\,2}$ szabadsági foka pedig 1.* □

3.2. állítás. (a) *Q_1, Q_2 és Q_3 függetlenek;*
(b) *Q_3/σ^2 mindig centrált χ^2-eloszlású $(I-1)(J-1)$ szabadsági fokkal;*
(c) *$\frac{1}{\sigma^2}Q_1 \sim \chi^2_{I-1}(\lambda_1)$, ahol $\lambda_1 = \frac{J}{\sigma^2}\sum_{i=1}^{I}\alpha_i^2$;*
(d) *$\frac{1}{\sigma^2}Q_2 \sim \chi^2_{J-1}(\lambda_2)$, ahol $\lambda_2 = \frac{I}{\sigma^2}\sum_{j=1}^{J}\beta_j^2$;*
(e) *$\frac{1}{\sigma^2}Q_1$ akkor és csak akkor centrált χ^2-eloszlású, ha H_A teljesül;*
(f) *$\frac{1}{\sigma^2}Q_2$ akkor és csak akkor centrált χ^2-eloszlású, ha H_B teljesül.*

Bizonyítás. Az előző állítás miatt a Fisher-Cochran-tétel feltétele teljesül, hiszen a rangok összege $(I-1)+(J-1)+(I-1)(J-1)+1 = IJ$. Ezért a kvadratikus formák függetlenek és χ^2-eloszlásúak. A nem-centralitási paramétereket úgy kapjuk, hogy a várható értékeket beírjuk a kvadratikus formákba. □

3.3. A szórásfelbontó táblázat

A 3.2. állítás alapján a Q_3 véletlen hiba (skalárszorosa) mindig centrált χ^2-eloszlású, Q_1 és Q_2 (skalárszorosa) viszont pontosan akkor, ha H_A, ill. H_B teljesül.

3.3. állítás. *Az*

$$F_A = \frac{Q_1}{I-1} \Big/ \frac{Q_3}{(I-1)(J-1)}$$

statisztika pontosan akkor $I-1$ és $(I-1)(J-1)$ szabadsági fokú centrált F-eloszlású, ha a

$$H_A : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

nullhipotézis teljesül.

Az

$$F_B = \frac{Q_2}{J-1} \Big/ \frac{Q_3}{(I-1)(J-1)}$$

statisztika pontosan akkor $J-1$ és $(I-1)(J-1)$ szabadsági fokú centrált F-eloszlású, ha a

$$H_B : \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

nullhipotézis teljesül.

BIZONYÍTÁS. A 3.2. állítás szerint F_A két független, a szabadsági fokával elosztott, (H_0 esetén centrált) χ^2-eloszlású valószínűségi változó hányadosa. Ezért az F-eloszlás definíciója alapján ennek eloszlása (H_0 esetén centrált) F-eloszlás $I-1$ és $(I-1)(J-1)$ szabadsági fokokkal. □

A fentiek alapján a kéttényezős (interakció nélküli) modell szórásfelbontó táblázata az alábbi.

A szóródás forrásai	Szabadsági fokok	Négyzet-összegek	Négyzet-átlagok	$F-$ hányadosok
Sorok közötti eltérések (A tényező szintjei)	$I-1$	Q_1	$S_1 = \frac{Q_1}{I-1}$	$\frac{S_1}{S_3}$
Oszlopok közötti eltérések (B tényező szintjei)	$J-1$	Q_2	$S_2 = \frac{Q_2}{J-1}$	$\frac{S_2}{S_3}$
Véletlen hiba	$(I-1)(J-1)$	Q_3	$S_3 = \frac{Q_3}{(I-1)(J-1)}$	
Teljes	$IJ-1$	Q		

A kéttényezős (interakció nélküli) modell szórásfelbontó táblázata

A H_A nullhipotézist $1-\alpha$ szinten elvetjük, ha

$$S_1/S_3 \geq F_{[I-1,\,(I-1)(J-1);\,\alpha]},$$

míg a H_B nullhipotézist $1-\alpha$ szinten elvetjük, ha

$$S_2/S_3 \geq F_{[J-1,\,(I-1)(J-1);\,\alpha]},$$

ahol az $F_{[l,m;\alpha]}$ az (l,m) szabadsági fokú centrált F-eloszlás $1-\alpha$ szinthez tartozó (felső) kritikus értéke.

3.4. Konfidencia intervallum a várható értékre

A nullhipotézisek elutasítása esetén az egytényezős modellnél mondottakhoz hasonlóan becsülhetjük az ismeretlen α_i, β_j paramétereket, különbségeiket és a σ^2-et.

Mivel Q_3/σ^2 mindig centrált χ^2-eloszlású $(I-1)(J-1)$ szabadsági fokkal, ezért σ^2 torzítatlan becslése $S_3 = \frac{Q_3}{(I-1)(J-1)}$.

Belátható, hogy $\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot} \sim \mathcal{N}\left(\alpha_i - \alpha_{i'}, \frac{2\sigma^2}{J}\right)$. Ezért $1-\alpha$ megbízhatósági szintű konfidencia intervallum az $\alpha_i - \alpha_{i'}$ különbségre (ahol $i \neq i'$)

$$(\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot}) \mp t_{[(I-1)(J-1);\,\alpha/2]}\sqrt{2S_3/J}, \tag{3.3}$$

ahol $t_{[(I-1)(J-1);\,\alpha/2]}$ az $1-\alpha$ szintű kétoldali t-próba kritikus értéke.

Továbbá, $\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$ eloszlása

$$\mathcal{N}\left(\alpha_i, \frac{\sigma^2}{J}\left(1 - \frac{1}{I}\right)\right), \tag{3.4}$$

ahol a szórás a lineáris forma együtthatói alapján adódott. Ezért az $1-\alpha$ szintű konfidencia intervallum α_i-re a következő:

$$(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}) \mp t_{[(I-1)(J-1);\,\alpha/2]}\sqrt{\frac{S_3}{J}\left(1 - \frac{1}{I}\right)}.$$

Hasonló módon adhatunk konfidencia intervallumot a $\beta_j - \beta_{j'}$ különbségre és β_j-re.

3.5. Feladatok

1. A (3.1)-ben szereplő mennyiségekkel definiáljuk az alábbi új mennyiségeket. $\widetilde{\mu} = \mu + \overline{\alpha} + \overline{\beta}$, $\widetilde{\alpha}_i = \alpha_i - \overline{\alpha}$, $\widetilde{\beta}_j = \beta_j - \overline{\beta}$, ahol $\overline{\alpha} = (1/I)\sum_{i=1}^{I}\alpha_i$, $\overline{\beta} = (1/J)\sum_{j=1}^{J}\beta_j$. Lássuk be, hogy minden i és j esetén $\widetilde{\mu} + \widetilde{\alpha}_i + \widetilde{\beta}_j = \mu + \alpha_i + \beta_j$, és a $\widetilde{\ }$-mal jelölt új mennyiségek már teljesítik a (3.2) feltételeket!

2. Lássuk be, hogy a (3.1) modell esetén, ha $i \neq i'$,

$$\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot} \sim \mathcal{N}\left(\alpha_i - \alpha_{i'}, \frac{2\sigma^2}{J}\right),$$

$\frac{1}{\sigma^2}Q_3 \sim \chi^2_{(I-1)(J-1)}$, és ezek egymástól függetlenek. Ez alapján igazoljuk (3.3)-at!

3. Igazoljuk (3.4)-et!

4. A kétszeres osztályozás az interakció figyelembevételével

4.1. A modell

Most a kétszeres osztályozás realisztikusabb esetét tekintjük: kétszeres osztályozást interakcióval.

Legyen az A tényezőnek I szintje, a B tényezőnek J szintje. Tételezzük fel tehát azt, hogy a két tényező együttes hatása (interakciója) is befolyásolhatja a kísérlet eredményét. Ilyenkor cellánként több megfigyelést kell végeznünk.

Végezzünk a két tényező IJ számú szintkombinációjának mindegyikére $K(>1)$ számú független megfigyelést. A cellánkénti egyenlő számú megfigyeléses kísérleti elrendezést **kiegyensúlyozott elrendezés**nek nevezzük. K-t ismétlésszámnak mondjuk.

Jelölje Y_{ijk} az (i,j)-edik cellában a k-adik megfigyelés eredményét $(k=1,\dots,K)$. Tegyük fel, hogy

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \tag{4.1}$$

ahol az ε_{ijk} véletlen hibák független, 0 és σ^2 paraméterű normális eloszlású valószínűségi változók.

A vizsgálandó hipotézisek:

$$\begin{aligned} H_A &: \alpha_1 = \alpha_2 = \dots = \alpha_I = 0\,, \\ H_B &: \beta_1 = \beta_2 = \dots = \beta_J = 0\,, \\ H_{AB} &: \text{minden } \ i=1,2,\dots,I,\ j=1,2,\dots,J\text{-re } \ \gamma_{ij}=0\,. \end{aligned}$$

Azaz nincs A-hatás, nincs B-hatás, illetve nincs AB interakció.

Elérhető, hogy a paraméterek kielégítsék a következő feltételeket:

$$\begin{aligned} &\sum_{i=1}^{I} \alpha_i = 0, \qquad \sum_{j=1}^{J} \beta_j = 0, \\ &\sum_{i=1}^{I} \gamma_{ij} = 0 \quad (j=1,2,\dots,J), \quad \sum_{j=1}^{J} \gamma_{ij} = 0 \quad (i=1,2,\dots,I). \end{aligned} \tag{4.2}$$

Ekkor az ismeretlen μ, α_i, β_j és γ_{ij} paraméterek legkisebb négyzetes becslései az alábbi alakúak. (A maximum likelihood becslések is ugyanezek.)

$$\begin{aligned} \widehat{\mu} &= \overline{Y}_{\dots}\,, \\ \widehat{\alpha}_i &= \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\dots}\,, \\ \widehat{\beta}_j &= \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\dots}\,, \\ \widehat{\gamma}_{ij} &= \overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\dots}\,, \end{aligned} \tag{4.3}$$

ahol

$$\overline{Y}_{...} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}\,,$$

$$\overline{Y}_{i..} = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}\,,$$

$$\overline{Y}_{.j.} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} Y_{ijk}\,,$$

$$\overline{Y}_{ij.} = \frac{1}{K}\sum_{k=1}^{K} Y_{ijk}$$

minden $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$ esetén. Ezek segítségével adhatjuk meg teljes négyzetösszeg felbontását. Az elemzés az alábbi négyzetösszegeken fog alapulni.

$$Q_1 = JK\sum_{i=1}^{I}(\overline{Y}_{i..} - \overline{Y}_{...})^2,$$

$$Q_2 = IK\sum_{j=1}^{J}(\overline{Y}_{.j.} - \overline{Y}_{...})^2,$$

$$Q_3 = K\sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2,$$

$$Q_4 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{ij.})^2,$$

$$Q = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{...})^2.$$

Q_1 az első tényező szintjei közötti eltéréseket, Q_2 a második tényező szintjei közötti eltéréseket, Q_3 pedig a két tényező együttes hatását írja le. Q_4 a véletlen hibák négyzetösszege, míg Q a teljes négyzetösszeg. A teljes négyzetösszeg felbontása:

$$Q = Q_1 + Q_2 + Q_3 + Q_4.$$

Ez a formula a kétszeres szorzatok 0 volta miatt igaz.

4.2. A Fisher-Cochran-tétel alkalmazása

A Fisher-Cochran-tétel alkalmazása szempontjából praktikusabb az alábbi felbontásból kiindulni.

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{ijk}^2 = Q_1 + Q_2 + Q_3 + Q_4 + IJK\,\overline{Y}_{...}^2\,. \tag{4.4}$$

Q_1 szabadsági foka $I-1$, Q_2 szabadsági foka $J-1$, Q_3 szabadsági foka $(I-1)(J-1)$, Q_4 szabadsági foka $IJ(K-1)$, $IJK\,\overline{Y}_{\cdots}^2$ szabadsági foka pedig 1.

4.1. állítás. *Az (4.1) modell esetén*

(a) Q_1, Q_2, Q_3 *és* Q_4 *függetlenek és* χ^2*-eloszlásúak, de nem feltétlenül centráltak;*
(b) $\frac{1}{\sigma^2}Q_4$ *mindig centrált* χ^2*-eloszlású* $IJ(K-1)$ *szabadsági fokkal;*
(c) $\frac{1}{\sigma^2}Q_1$ *akkor és csak akkor centrált* χ^2_{I-1}*-eloszlású, ha* H_A *teljesül;*
(d) $\frac{1}{\sigma^2}Q_2$ *akkor és csak akkor centrált* χ^2_{J-1}*-eloszlású, ha* H_B *teljesül;*
(e) $\frac{1}{\sigma^2}Q_3$ *akkor és csak akkor centrált* $\chi^2_{(I-1)(J-1)}$*-eloszlású, ha* H_{AB} *teljesül.*

BIZONYÍTÁS. A (4.4) előállításban a Fisher-Cochran-tétel feltétele teljesül, hiszen a rangok összege $(I-1)+(J-1)+(I-1)(J-1)+IJ(K-1)+1=IJK$. Ezért a kvadratikus formák függetlenek és χ^2-eloszlásúak. A nem-centralitási paramétereket úgy kapjuk, hogy a várható értékeket beírjuk a kvadratikus formákba. □

4.3. A szórásfelbontó táblázat

A fentiek alapján a kéttényezős (interakciót figyelembe vevő) modell szórásfelbontó táblázata az alábbi.

A szóródás forrásai	Szabadsági fokok	Négyzet-összegek	Négyzet-átlagok	F-hányadosok
A tényező	$I-1$	Q_1	$S_1=\frac{Q_1}{I-1}$	S_1/S_4
B tényező	$J-1$	Q_2	$S_2=\frac{Q_2}{J-1}$	S_2/S_4
AB interakció	$(I-1)(J-1)$	Q_3	$S_3=\frac{Q_3}{(I-1)(J-1)}$	S_3/S_4
Véletlen hiba	$IJ(K-1)$	Q_4	$S_4=\frac{Q_4}{IJ(K-1)}$	
Teljes	$IJK-1$	Q		

Szórásfelbontó táblázat kéttényezős (interakciót figyelembe vevő) modellre

A különböző nullhipotézisek F-próbáit ugyanúgy hajtjuk végre, mint az előző modelleknél. Elsőként a H_{AB} hipotézist vizsgáljuk.

4.1. példa. Kétféle vas (Fe^{2+} és Fe^{3+}) szervezetből való kiürülését vizsgáljuk. Mindkét vas fajtából 10.2, 1.2 és 0.3 millimólos koncentrációt adunk, 18-18 egyednek.

A lenti mátrix első 3 oszlopa a Fe^{2+} három különböző mennyisége esetén a szervezetben maradt százalékos arányt mutatja a 18-18 egyed esetén. Hasonló a mátrix további 3 oszlopa. A 3 különböző vas-mennyiség az oszlophatás. A vas 2 minősége a sorhatás. Végül 18 az ismétlésszám. Tehát egy kiegyensúlyozott elrendezésűnk van.

Miért nem végeztük el mégsem az ANOVA eljárást ezekre az adatokra? Mert a mintának normális eloszlásúnak kell lennie azonos szórással! Viszont az empirikus szórásnégyzetek az mutatták, hogy cellánként vett 18-18 elemű minta szórásai nagyon eltérnek. Ugyanezen mintákra a hisztogram a haranggörbétől nagy eltérés mutatott, ugyanígy a Gauss-papír az egyenestől, tehát a normalitás sem áll fenn. Próbálkozzunk a minta transzformálásával! A logaritmus vétel jelentős javulást hozott mind a normalitásban, mind a szórások egyenlőségében. (Megjegyezzük, hogy normalitást 18 elemű

mintából nem igazán lehet jól tesztelni. A módszert tekintve: a normalitást Shapiro-Wilk-próbával, a szórások azonosságát egy Bartlett-próbával kellene ellenőrizni.)

Fe^{2+}			Fe^{3+}		
10.2	1.2	0.3	10.2	1.2	0.3
0.71	2.20	2.25	2.20	4.04	2.71
1.66	2.93	3.93	2.69	4.16	5.43
2.01	3.08	5.08	3.54	4.42	6.38
2.16	3.49	5.82	3.75	4.93	6.38
2.42	4.11	5.84	3.83	5.49	8.32
2.42	4.95	6.89	4.08	5.77	9.04
2.56	5.16	8.50	4.27	5.86	9.56
2.60	5.54	8.56	4.53	6.28	10.01
3.31	5.68	9.44	5.32	6.97	10.08
3.64	6.25	10.52	6.18	7.06	10.62
3.74	7.25	13.46	6.22	7.78	13.80
3.74	7.90	13.57	6.33	9.23	15.99
4.39	8.85	14.56	6.97	9.34	17.90
4.50	11.96	16.41	6.97	9.91	18.25
5.07	15.54	16.96	7.52	13.46	19.32
5.26	15.89	17.56	8.36	18.4	19.87
8.15	18.3	22.82	11.65	23.89	21.60
8.24	18.59	29.13	12.45	26.39	22.25

Az adatok logaritmusára az alábbi szorásfelbontó táblázatot kaptuk:

ANOVA Table				
Source	SS	df	MS	F
Columns	15.58	2	7.79	22.52
Rows	2.079	1	2.079	6.01
Interaction	0.8089	2	0.4045	1.62
Error	35.39	102	0.346	
Total	53.76	107		

A szokásos H_A, H_B, H_{AB} nullhipotézisek esetén a megfelelő F-próbák pontos terjedelmére a következőket kaptuk: $7.9 * 10^{-9}$, 0.015, 0.314. Ezen p értékek alapján el kell vetni, hogy H_A: *nincs oszlophatás*, H_B: *nincs sorhatás*, de el kell fogadni, hogy H_{AB}: *nincs interakció*.

Magyarázat. SS oszlopában a teljes négyzetösszeg felbontása van oszlophatásra, sorhatásra, interakcióra és véletlen hibára. A df alatt ezek szabadsági foka áll. MS alatt SS és df hányadosa. F pedig a megfelelő hatás MS értékének és a véletlen hiba MS értékének hányadosa. Ha valamelyik hatás a véletlen hibához képest túl nagy, akkor elvetjük az azon hatásra vonatkozó nullhipotézist. A $(p_1, p_2, p_3) = (7.9 * 10^{-9}, 0.015, 0.314)$ vektor i-edik komponense annak az értékét adja, hogy a nullhipotézis igaz volta esetén milyen valószínűséggel lehetne a centrált F-eloszlású valószínűségi

változó legalább olyan nagy, mint a kapott F érték. Így kicsi p_i esetén elvetjük a nullhipotézist. A p_i értékét F-eloszlás alapján határozzuk meg: az F oszlopban lévő statisztikák mindig F-eloszlásúak, a megfelelő nullhipotézis fennállása esetén centráltak, ha viszont a nullhipotézis nem igaz, akkor nem centráltak. (Megjegyezzük, hogy a két fenti szórásanalízisbeli példában a MATLAB és a SAS futási eredménye teljesen megegyezett.) □

4.4. Feladatok

1. A (4.1)-ben szereplő mennyiségekkel definiáljuk az alábbi új mennyiségeket. $\widetilde{\mu} = \mu + \overline{\alpha} + \overline{\beta} + \overline{\gamma}_{\cdot\cdot}$, $\widetilde{\alpha}_i = \alpha_i - \overline{\alpha} + \overline{\gamma}_{i\cdot} - \overline{\gamma}_{\cdot\cdot}$, $\widetilde{\beta}_j = \beta_j - \overline{\beta} + \overline{\gamma}_{\cdot j} - \overline{\gamma}_{\cdot\cdot}$, $\widetilde{\gamma}_{ij} = \gamma_{ij} - \overline{\gamma}_{i\cdot} - \overline{\gamma}_{\cdot j} + \overline{\gamma}_{\cdot\cdot}$, ahol $\overline{\alpha} = (1/I)\sum_{i=1}^{I}\alpha_i$, $\overline{\beta} = (1/J)\sum_{j=1}^{J}\beta_j$, $\overline{\gamma}_{i\cdot} = (1/J)\sum_{j=1}^{J}\gamma_{ij}$, $\overline{\gamma}_{\cdot j} = (1/I)\sum_{i=1}^{I}\gamma_{ij}$, $\overline{\gamma}_{\cdot\cdot} = (1/IJ)\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_{ij}$. Lássuk be, hogy minden i és j esetén $\widetilde{\mu} + \widetilde{\alpha}_i + \widetilde{\beta}_j + \widetilde{\gamma}_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, és a $\widetilde{\ }$-mal jelölt új mennyiségek már teljesítik a (4.2) feltételeket!

2. A (4.1) modellben a legkisebb négyzetes becslés, azaz a hiba négyzetösszeg minimalizálása a

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\Big(Y_{ijk} - (\mu + \alpha_i + \beta_j + \gamma_{ij})\Big)^2 \tag{4.5}$$

függvény minimumának meghatározását jelenti μ, α_i, β_j és γ_{ij} szerint. Lássuk be, hogy a minimum a (4.3) képletbeni értékeknél adódik! Ehhez használjuk a minimum deriváltakkal történő meghatározását. A kapott lineáris egyenletrendszer megoldásakor vegyük figyelembe (4.2)-ot.

II. fejezet
Regresszióanalízis és a lineáris modell

1. A lineáris modell

1.1. A lineáris modell definíciója

(1.1) $$\boxed{\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$$

a lineáris modell, ahol

$\boldsymbol{Y}$ n-dimenziós megfigyelés vektor,
X $n \times p$ méretű, nem véletlen, megfigyelt mátrix (a magyarázó változók mátrixa),
$\boldsymbol{\beta}$ p-dimenziós ismeretlen paraméter,
$\boldsymbol{\varepsilon}$ nem megfigyelhető n-dimenziós véletlen vektor (hiba).

Általában $n \gg p$, ezt szükség esetén fel fogjuk tenni. A gyakorlatban p a magyarázó változók száma, n pedig a megfigyelt objektumok száma, tehát $n \gg p$ ésszerű feltétel.

1.1. példa.

1.2. A legkisebb négyzetek módszere

Ha $\mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}$ és $\operatorname{var}\boldsymbol{\varepsilon} = \sigma^2 I$ (σ^2 ismeretlen paraméter), akkor **homoszkedasztikus** esetről beszélünk. Ekkor a **legkisebb négyzetes becslést** (OLS=Ordinary Least Squares) alkalmazzuk $\boldsymbol{\beta}$-ra: ez lesz $\widehat{\boldsymbol{\beta}}$.

Legyen tehát $\widehat{\boldsymbol{\beta}}$ az $\|\boldsymbol{Y} - X\boldsymbol{\beta}\|^2$-et minimalizáló vektor. (Itt $\|\,.\,\|$ a norma $\mathbb{R}^n$-ben.) Jelölje P_F az X oszlopai által generált F altérre való merőleges vetítést.

1.1. állítás. 1. *$\widehat{\boldsymbol{\beta}}$ legkisebb négyzetes becslés* $\iff$

$$X\widehat{\boldsymbol{\beta}} = P_F\boldsymbol{Y}.$$

2. *$\widehat{\boldsymbol{\beta}}$ legkisebb négyzetes becslés* $\iff$ *$\widehat{\boldsymbol{\beta}}$ az*

$$\boxed{X^\top\boldsymbol{Y} = X^\top X\boldsymbol{\beta}}$$

normálegyenlet megoldása.

Bizonyítás. **1.** $\boldsymbol{Y}$-hoz az X oszlopai által generált altér mely $X\boldsymbol{\beta}$ eleme lesz legközelebb? Éppen az $\boldsymbol{Y}$ vetülete, azaz $P_F\boldsymbol{Y}$. Így $X\widehat{\boldsymbol{\beta}} = P_F\boldsymbol{Y}$.

2. $\|\boldsymbol{Y} - X\boldsymbol{\beta}\|^2$ mikor a legkisebb? Ha $\boldsymbol{Y} - X\boldsymbol{\beta}$ éppen az $\boldsymbol{Y}$ ortogonális komplementere az F altérre vonatkozóan. Azaz $\boldsymbol{Y} - X\boldsymbol{\beta}$ merőleges X minden oszlopára, tehát

$$X^\top\boldsymbol{Y} - X^\top X\boldsymbol{\beta} = \mathbf{0},$$

vagyis

$$X^\top X\widehat{\boldsymbol{\beta}} = X^\top\boldsymbol{Y}. \qquad \square$$

1.1. megjegyzés. $X^\top X$ invertálható $\iff$ $\operatorname{rang} X = p$.

BIZONYÍTÁS. **1.** Ha $X^\top X$ invertálható, akkor $X^\top X\boldsymbol{a} = \mathbf{0}$ esetén $\boldsymbol{a}$=$\mathbf{0}$. Ezért $X\boldsymbol{a}$ is csak akkor lehet $\mathbf{0}$, ha $\boldsymbol{a} = \mathbf{0}$. Azaz X oszlopainak csak 0 együtthatókkal való lineáris kombinációi egyenlőek $\mathbf{0}$-val, tehát X rangja egyenlő az oszlopai számával, vagyis p-vel.

2. Ha $\operatorname{rang} X = p$, akkor $X\boldsymbol{a} \neq \mathbf{0}$ tetszőleges $\boldsymbol{a} \neq \mathbf{0}$ esetén; így $\boldsymbol{a}^\top X^\top X\boldsymbol{a} = \|X\boldsymbol{a}\|^2 \neq 0$. Viszont ha $X^\top X$ nem lenne invertálható, akkor a vele képzett kvadratikus forma 0 lenne valamely $\boldsymbol{a} \neq \mathbf{0}$ vektorra. □

1.2. megjegyzés. Ha $\operatorname{rang} X = p$, akkor

$$\boxed{\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{Y}}.$$

Ez éppen a normálegyenlettel ekvivalens, ha $(X^\top X)$ invertálható. □

1.2. állítás. *Legyen* $\mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}$, $\operatorname{var}\boldsymbol{\varepsilon} = \sigma^2 I$ *és* $\operatorname{rang} X = p$. *Ekkor* $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{Y}$ *torzítatlan becslése* $\boldsymbol{\beta}$*-nak, továbbá* $\operatorname{var}\widehat{\boldsymbol{\beta}} = \sigma^2 (X^\top X)^{-1}$.

Ha $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 I)$, *akkor* $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1})$.

BIZONYÍTÁS. Ha $\operatorname{rang} X = p$, akkor $(X^\top X)$ invertálható. Ekkor

$$\mathbb{E}\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbb{E}\boldsymbol{Y} = \boldsymbol{\beta},$$

hiszen $\mathbb{E}\boldsymbol{Y} = X\boldsymbol{\beta}$. Másrészt

$$\operatorname{var}(\widehat{\boldsymbol{\beta}}) = (X^\top X)^{-1} X^\top (\operatorname{var}(\boldsymbol{Y})) X (X^\top X)^{-1} = \sigma^2 (X^\top X)^{-1},$$

ugyanis $\operatorname{var}(\boldsymbol{Y}) = \operatorname{var}(\boldsymbol{\varepsilon}) = \sigma^2 I$.

Ha $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 I)$, akkor $\boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I)$, így $\widehat{\boldsymbol{\beta}}$ — lévén $\boldsymbol{Y}$ lineáris függvénye — maga is normális eloszlású. □

1.3. A Gauss-Markov-tétel legegyszerűbb alakja

A homoszkedasztikus esetben $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{Y}$ legkisebb négyzetes becslés a **legjobb lineáris torzítatlan becslés** (BLUE=Best Linear Unbiased Estimator). Ezt mondja ki a **Gauss-Markov-tétel** legegyszerűbb alakja.

1.1. tétel. (Gauss-Markov.) *Ha* $\mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}$, $\operatorname{var}\boldsymbol{\varepsilon} = \sigma^2 I$ *és* $\operatorname{rang} X = p$, *akkor* $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{Y}$ *a* $\boldsymbol{\beta}$ *paraméter vektor legjobb lineáris torzítatlan becslése.*

BIZONYÍTÁS. $\widehat{\boldsymbol{\beta}}$ az $\boldsymbol{Y}$ vektor lineáris függvénye ($(X^\top X)^{-1} X^\top$ mátrixszal való szorzata), azaz $\widehat{\boldsymbol{\beta}}$ lineáris becslés. Továbbá $\mathbb{E}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, tehát torzítatlan.

Legyen $C\boldsymbol{Y}$ a $\boldsymbol{\beta}$-nak egy másik lineáris torzítatlan becslése. Be kell látnunk, hogy $C\boldsymbol{Y}$ szórásmátrixa „nagyobb” $\widehat{\boldsymbol{\beta}}$ szórásmátrixánál, azaz $\operatorname{var}(C\boldsymbol{Y}) - \operatorname{var}(\widehat{\boldsymbol{\beta}})$ pozitív szemidefinit. $C\boldsymbol{Y}$ torzítatlanságából

$$\boldsymbol{\beta} = \mathbb{E}C\boldsymbol{Y} = C\mathbb{E}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = CX\boldsymbol{\beta} \qquad \forall\, \boldsymbol{\beta}\,.$$

Tehát $CX = I$.

$$\begin{aligned}\operatorname{var}(C\boldsymbol{Y}) &= \mathbb{E}(C\boldsymbol{Y}-\boldsymbol{\beta})(C\boldsymbol{Y}-\boldsymbol{\beta})^\top = \\ &= \mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}}+\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}}+\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top = \\ &= \mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})^\top + \mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top + \mathbb{E}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})^\top \\ &+ \mathbb{E}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \geq \mathbb{E}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top = \operatorname{var}(\widehat{\boldsymbol{\beta}}),\end{aligned}$$

hiszen

$$\mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})^\top \qquad \text{pozitív szemidefinit, és}$$

$$\mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top = 0$$

(ezt utóbbit alább megmutatjuk). Így $\operatorname{var}(C\boldsymbol{Y}) \geq \operatorname{var}(\widehat{\boldsymbol{\beta}})$, azaz $\widehat{\boldsymbol{\beta}}$ a legjobb (a legkisebb szórású).

Figyeljük meg, hogy

$$\begin{aligned}C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}} &= \underbrace{CX}_{I}\boldsymbol{\beta} + C\boldsymbol{\varepsilon} - (X^\top X)^{-1}X^\top X\boldsymbol{\beta} - (X^\top X)^{-1}X^\top\boldsymbol{\varepsilon} \\ &= [C-(X^\top X)^{-1}X^\top]\boldsymbol{\varepsilon}\end{aligned}$$

és $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} = (X^\top X)^{-1}X^\top\boldsymbol{\varepsilon}$, ahonnan

$$\begin{aligned}\mathbb{E}(C\boldsymbol{Y}-\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top &= [C-(X^\top X)^{-1}X^\top]\underbrace{\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)}_{\sigma^2 I}X(X^\top X)^{-1} = \\ &= \sigma^2\big[\underbrace{CX}_{I}(X^\top X)^{-1} - \underbrace{(X^\top X)^{-1}X^\top X}_{I}(X^\top X)^{-1}\big] = 0. \qquad \square\end{aligned}$$

1.4. Mátrixok általánosított inverze

1.1. definíció. Az A $n \times n$-es mátrix **általánosított inverze** az az A^- $n \times n$-es mátrix, melyre

$$\boxed{AA^-A = A}. \qquad \square$$

1.3. megjegyzés. Ha A invertálható, akkor A^- egyértelműen létezik, és $A^- = A^{-1}$.

Ugyanis ekkor $AA^{-1}A = A$, azaz A^{-1} kielégíti A^- definícióját. Másrészt az $AA^-A = A$ egyenletet jobbról és balról is megszorozva A^{-1}-gyel, $A^- = A^{-1}$-et kapjuk. $\square$

1.3. állítás. *Ha A szimmetrikus, akkor létezik általánosított inverze.*

Bizonyítás. Legyenek $\lambda_1, \ldots, \lambda_k$ az A nem zérus sajátértékei, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ pedig az A sajátvektorainak ortonormált rendszere. Legyen U az $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$-et tartalmazó ortogonális mátrix,

$$\Lambda^- = \begin{pmatrix} \lambda_1^{-1} & & & & \\ & \ddots & & & \\ & & \lambda_k^{-1} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_k & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

$n \times n$-es diagonális mátrixok. Ekkor $A^- = U\Lambda^- U^\top$ teljesíti a feltételeket:

$$AA^-A = (U\Lambda U^\top)(U\Lambda^- U^\top)(U\Lambda U^\top) = U\Lambda U^\top = A. \qquad \square$$

1.4. megjegyzés. Erre a speciális A^--ra $A^-AA^- = A^-$ és $(A^-)^\top = A^-$ is teljesül. $\square$

1.4. állítás. *$X(X^\top X)^- X^\top$ ortogonális projekció az X oszlopvektorai által generált F altérre.*

Bizonyítás. Ha $\boldsymbol{w} \perp F$, akkor $X^\top \boldsymbol{w} = \boldsymbol{0}$, így $X(X^\top X)^- X^\top \boldsymbol{w} = \boldsymbol{0}$.

Ahhoz, hogy $X(X^\top X)^- X^\top$ az F-beli vetorokat önmagukba viszi, elég belátni, hogy $X(X^\top X)^- X^\top X = X$. Ez viszont igaz, mert tetszőleges $\boldsymbol{v}$ vektor $\boldsymbol{v} = \boldsymbol{v}_1 + X\boldsymbol{v}_2$ alakba írható, ahol $\boldsymbol{v}_1 \perp F$. Ezért

$$\begin{aligned} \boldsymbol{v}^\top X(X^\top X)^- X^\top X &= \boldsymbol{v}_1^\top X(X^\top X)^- X^\top X + \boldsymbol{v}_2^\top X^\top X(X^\top X)^- X^\top X = \\ &= \boldsymbol{0} + \boldsymbol{v}_2^\top (X^\top X) = \boldsymbol{v}^\top X. \end{aligned} \qquad \square$$

1.5. A Gauss-Markov-tétel általános alakja

1.5. megjegyzés. $\widehat{\boldsymbol{\beta}} = (X^\top X)^- X^\top \boldsymbol{Y}$ mindig. Hiszen ekkor az $X\widehat{\boldsymbol{\beta}} = P_F \boldsymbol{Y}$ egyenlet teljesül. Ugyanis $X(X^\top X)^- X^\top$ éppen a P_F vetítéssel egyenlő. $\square$

1.6. megjegyzés. Ha X oszlopvektorai lineárisan függőek, akkor azt mondjuk, hogy a magyarázó változók között **kollinearitás** áll fenn. Ekkor — azaz ha $\operatorname{rang} X < p$ — a $\boldsymbol{\beta}$ vektornak csupán bizonyos koordinátáit, illetve azok lineáris függvényeit tudjuk becsülni, a teljes $\boldsymbol{\beta}$-t nem. $\square$

1.2. definíció. A $\boldsymbol{\beta}$ paraméter $\boldsymbol{c}^\top \boldsymbol{\beta}$ lineáris függvényét (lineárisan és torzítatlanul) **becsülhetőnek** nevezzük, ha van $\boldsymbol{Y}$-nak olyan $\boldsymbol{b}^\top \boldsymbol{Y}$ lineáris függvénye, melyre

$$\mathbb{E}\boldsymbol{b}^\top \boldsymbol{Y} = \boldsymbol{c}^\top \boldsymbol{\beta} \qquad \forall \boldsymbol{\beta} \in \mathbb{R}^p.$$

(Itt $\boldsymbol{c} \in \mathbb{R}^p$, $\boldsymbol{b} \in \mathbb{R}^n$.) Ilyenkor röviden azt is mondjuk, hogy $\boldsymbol{c}^\top$ becsülhető. $\square$

Amikor $\operatorname{rang} X = p$, akkor a teljes $\boldsymbol{\beta}$ vektor becsülhető.

Az előzőekben az, hogy $\widehat{\boldsymbol{\beta}}$ a legjobb lineáris torzítatlan becslés, azt jelentette, hogy $\widehat{\boldsymbol{\beta}}$ szórásmátrixa a legkisebb a lineáris torzítatlan becslések szórásmátrixai között. (A rendezést a „$D_1 \geq D_2 \iff D_1 - D_2$ pozitív szemidefinit” alapján éretelmezzük.) Tehát $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$ szórása a legkisebb a $\boldsymbol{c}^\top \boldsymbol{\beta}$ lineáris torzítatlan becslései szórásai között ($\boldsymbol{c} \in \mathbb{R}^p$). Ez a tény a Gauss-Markov-tétel általánosabb változatának is a lényege.

1.2. tétel. (Gauss-Markov.) **1.** *$\boldsymbol{c}^\top \boldsymbol{\beta}$ becsülhető $\iff$ $\boldsymbol{c} = X^\top \boldsymbol{b}$ valamely $\boldsymbol{b} \in \mathbb{R}^n$-re.*

2. *A $\boldsymbol{c}^\top \boldsymbol{\beta}$ minden lineáris torzítatlan $\boldsymbol{b}^\top Y$ becslésére $P_F \boldsymbol{b} = \boldsymbol{b}^*$, ahol $\boldsymbol{b}^*$ (a $\boldsymbol{b}$-től nem függő) állandó vektor. Itt $P_F = X(X^\top X)^- X^\top$ az X oszlopai alterére való vetítés.*

3. *$\boldsymbol{b}^{*\top} \boldsymbol{Y}$ adja a legkisebb szórású lineáris torzítatlan becslést.*

4. *$\boldsymbol{b}^{*\top} Y = \boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$.*

BIZONYÍTÁS. **1.** Legyen $\boldsymbol{c}^\top\boldsymbol{\beta}$ becsülhető. Ekkor $\boldsymbol{c}^\top\boldsymbol{\beta} = \mathbb{E}(\boldsymbol{b}^\top\boldsymbol{Y})$ valamely $\boldsymbol{b}$-re. Azaz $\boldsymbol{c}^\top\boldsymbol{\beta} = \boldsymbol{b}^\top X\boldsymbol{\beta}\ \ \forall\boldsymbol{\beta}$. Tehát $\boldsymbol{c}^\top = \boldsymbol{b}^\top X$.

Megfordítva, $\boldsymbol{c}^\top\boldsymbol{\beta} = \boldsymbol{b}^\top X\boldsymbol{\beta}$ becsülhető $\boldsymbol{b}^\top\boldsymbol{Y}$-nal.

2. Legyen $\boldsymbol{b}^\top\boldsymbol{Y}$ a $\boldsymbol{c}^\top\boldsymbol{\beta}$ lineáris torzítatlan becslése. Ekkor tekintsük a

$$\boldsymbol{b} = (\boldsymbol{b} - P_F\boldsymbol{b}) + P_F\boldsymbol{b}$$

ortogonális felbontást. Ezt $\boldsymbol{Y}$-nal jobbról szorozva és várható értéket véve:

$$\mathbb{E}\boldsymbol{b}^\top\boldsymbol{Y} = (\boldsymbol{b} - P_F\boldsymbol{b})^\top X\boldsymbol{\beta} + \mathbb{E}(P_F\boldsymbol{b})^\top\boldsymbol{Y}.$$

Mivel $\boldsymbol{b} - P_F\boldsymbol{b}$ és $X\boldsymbol{\beta}$ egymásra merőleges és $\mathbb{E}\boldsymbol{b}^\top\boldsymbol{Y} = \boldsymbol{c}^\top\boldsymbol{\beta}$, így

$$\boldsymbol{c}^\top\boldsymbol{\beta} = \mathbb{E}\boldsymbol{b}^{*\top}\boldsymbol{Y}$$

adódik, ahol $\boldsymbol{b}^* = P_F\boldsymbol{b}$.

Legyen most $\boldsymbol{b}_1^\top\boldsymbol{Y}$ a $\boldsymbol{c}^\top\boldsymbol{\beta}$ egy másik lineáris torzítatlan becslése. Ekkor

$$0 = \boldsymbol{c}^\top\boldsymbol{\beta} - \boldsymbol{c}^\top\boldsymbol{\beta} = \mathbb{E}\boldsymbol{b}^{*\top}\boldsymbol{Y} - \mathbb{E}\boldsymbol{b}_1^\top\boldsymbol{Y} = \left(\boldsymbol{b}^{*\top} - \boldsymbol{b}_1^\top\right)X\boldsymbol{\beta} \qquad \forall\boldsymbol{\beta}.$$

Azaz $\boldsymbol{b}^* - \boldsymbol{b}_1$ merőleges X oszlopaira. Tehát $\boldsymbol{b}_1 = \boldsymbol{b}^* + (\boldsymbol{b}_1 - \boldsymbol{b}^*)$ a merőleges felbontás, azaz $\boldsymbol{b}^*$ a $\boldsymbol{b}_1$-nek is vetülete.

3. Az előzőekben kaptuk, hogy $\mathbb{E}\boldsymbol{b}^{*\top}\boldsymbol{Y} = \boldsymbol{c}^\top\boldsymbol{\beta}$, azaz $\boldsymbol{b}^{*\top}\boldsymbol{Y}$ lineáris torzítatlan becslés. Az ortogonalitás alapján:

$$\mathbb{D}^2\boldsymbol{b}^\top\boldsymbol{Y} = \sigma^2\boldsymbol{b}^\top\boldsymbol{b} = \sigma^2\left[\boldsymbol{b}^{*\top}\boldsymbol{b}^* + (\boldsymbol{b} - \boldsymbol{b}^*)^\top(\boldsymbol{b} - \boldsymbol{b}^*)\right] \geq \sigma^2\boldsymbol{b}^{*\top}\boldsymbol{b}^* = \mathbb{D}^2(\boldsymbol{b}^{*\top}\boldsymbol{Y}).$$

Tehát $\boldsymbol{b}^{*\top}\boldsymbol{Y}$ szórása a legkisebb.

4. Mivel $X\widehat{\boldsymbol{\beta}} = P_F\boldsymbol{Y}$, így $\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}$ merőleges X oszlopai alterére. Viszont $\boldsymbol{b}^*$ benne van ebben az altérben. Tehát $\boldsymbol{b}^{*\top}(\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}) = 0$. Azaz

$$\boldsymbol{b}^{*\top}\boldsymbol{Y} = \boldsymbol{b}^{*\top}X\widehat{\boldsymbol{\beta}} = \boldsymbol{c}^\top\widehat{\boldsymbol{\beta}},$$

mert $\boldsymbol{c} = X^\top\boldsymbol{b}$ és $X^\top\boldsymbol{b} = X^\top\boldsymbol{b}^*$, lévén $\boldsymbol{b}^*$ a $\boldsymbol{b}$ vetülete X oszlopai alterére. □

2. A lineáris modell normális eloszlás esetén

2.1. Maximum-likelihood becslés

2.1. tétel. *Ha $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 I)$, akkor $\boldsymbol{\beta}$ maximum-likelihood becslése $\widehat{\boldsymbol{\beta}}$, σ^2 maximum-likelihood becslése pedig $\frac{1}{n}\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2$.*

Bizonyítás. Mivel $\boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I)$, így $\boldsymbol{Y}$ sűrűségfüggvénye:

$$f(\boldsymbol{y}) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|^2\right\}.$$

Ennek $\boldsymbol{\beta}$ szerinti maximuma az exponensben lévő kifejezés, nevezetesen $\|\boldsymbol{Y} - X\boldsymbol{\beta}\|^2$ minimuma helyén van, azaz $\widehat{\boldsymbol{\beta}}$-ban.

A σ^2 szerinti minimumához

$$0 = \frac{\partial \log f}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{(\sigma^2)^2}\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2.$$

Innen $\widehat{\sigma}^2 = \frac{1}{n}\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2$. □

2.1. megjegyzés. $\widehat{\sigma}^2 = \frac{1}{n}\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2$ a becsült $\boldsymbol{\beta}$ esetén a **maradék négyzetösszeg**, (RSS=Residual Sum of Squares) azaz $\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2$ n-ed része. □

2.1. lemma. *Legyen V egy s-dimenziós altérre való merőleges vetítés mátrixa, $\boldsymbol{\eta} \sim \mathcal{N}_n(\boldsymbol{m}, I)$. Ekkor $V\boldsymbol{\eta}$ és $(I - V)\boldsymbol{\eta}$ függetlenek és normális eloszlásúak. Továbbá*

$$\|V(\boldsymbol{\eta} - \boldsymbol{m})\|^2 \sim \chi^2_s.$$

Bizonyítás. Nyilván $\begin{pmatrix} V\boldsymbol{\eta} \\ (I-V)\boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} V \\ I - V \end{pmatrix}\boldsymbol{\eta}$, azaz normális eloszlású vektor lineáris transzformáltja, tehát normális eloszlású. Másrészt

$$\operatorname{cov}(V\boldsymbol{\eta}, (I-V)\boldsymbol{\eta}) = VI(I-V) = 0,$$

azaz korrelálatlanok, tehát függetlenek.

Legyen $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_s$ az s-dimenziós alterünk ortonormált bázisa. Ezt egészítsük ki ortonormált bázissá: $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Ezen vektorokat foglaljuk az U ortogonális mátrixba. Ekkor $V = U\Lambda U^\top$, ahol Λ olyan diagonális mátrix, melynek első s főátlós eleme 1, a többi 0.

$$\|V(\boldsymbol{\eta} - \boldsymbol{m})\|^2 = \|U\Lambda U^\top(\boldsymbol{\eta} - \boldsymbol{m})\|^2 = \|U\Lambda\boldsymbol{\xi}\|^2 = \boldsymbol{\xi}^\top\Lambda U^\top U\Lambda\boldsymbol{\xi},$$

ahol $\boldsymbol{\xi} \sim \mathcal{N}_n(\mathbf{0}, I)$. De $U^\top U = I$ és $\Lambda\Lambda = \Lambda$, tehát

$$\|V(\boldsymbol{\eta} - \boldsymbol{m})\|^2 = \xi_1^2 + \cdots + \xi_s^2 \sim \chi^2_s.$$ □

2.2. tétel. *Legyen $\boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I)$, $r = \operatorname{rang} X$. Ekkor*

$\frac{1}{\sigma^2}\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2 \sim \chi^2_{n-r}$,

$\frac{\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}\|^2}{n-r}$ a σ^2 paraméter torzítatlan becslése.

BIZONYÍTÁS. Emlékeztetünk, hogy F az X oszlopvektorai által kifeszített altér, P_F pedig az F-re merőleges vetítés.

$\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}} = \boldsymbol{Y}-P_F\boldsymbol{Y} = (I-P_F)\boldsymbol{Y}$. Mivel $I-P_F$ az F ortogonális komplementerére való merőleges vetítés, így $\mathbb{E}(\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}) = (I-P_F)\mathbb{E}\boldsymbol{Y} = (I-P_F)X\boldsymbol{\beta} = \boldsymbol{0}$. Tehát $\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2(I-P_F))$.

Tudjuk, hogy $(I-P_F)(\boldsymbol{Y}-\mathbb{E}\boldsymbol{Y}) = (I-P_F)\boldsymbol{Y} = \boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}$. Az 2.1. lemma miatt ezek közül a legelső hossznégyzetének $1/\sigma^2$-szerese χ^2_{n-r} eloszlású, így $\frac{1}{\sigma^2}\|\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}\|^2 \sim \chi^2_{n-r}$.

Mivel $\mathbb{E}\chi^2_{n-r} = n-r$, ezért $\mathbb{E}\frac{\|\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}\|^2}{n-r} = \sigma^2$. □

2.2. megjegyzés. $\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}$ és $X\widehat{\boldsymbol{\beta}}$ függetlenek.

Ugyanis $\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}} = (I-P_F)\boldsymbol{Y}$ és $X\widehat{\boldsymbol{\beta}} = P_F\boldsymbol{Y}$. □

2.2. Hipotézisvizsgálat a lineáris modellben

Legyen $Y = X\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, ahol $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 I)$. A nullhipotézis:

$$H_0 : B\boldsymbol{\beta} = \boldsymbol{0},$$

ahol a B mátrix ismert és sorai becsülhetőek.

2.3. megjegyzés. Legyen az X oszlopai által generált altér F. Legyenek B sorai becsülhetőek. Ekkor létezik olyan F_0 altere F-nek, hogy $B\boldsymbol{\beta} = \boldsymbol{0} \iff X\boldsymbol{\beta} \in F_0$.

Ennek bizonyítására megjegyezzük, hogy B sorai becsülhetőek $\iff B = B^*X$ valamely B^* mátrixra. Ezért $B\boldsymbol{\beta} = \boldsymbol{0} \iff B^*X\boldsymbol{\beta} = \boldsymbol{0} \iff X\boldsymbol{\beta}$ a B^* mátrix nullterében van, azaz $X\boldsymbol{\beta} \in F_0$, ahol F_0 a $B^*{}_{|F}$ nulltere. □

2.3. tétel. *Legyen*

$$S_B^2 = \min_{\boldsymbol{\beta}\,:\,B\boldsymbol{\beta}=\boldsymbol{0}} \|\boldsymbol{Y}-X\boldsymbol{\beta}\|^2, \qquad S_e^2 = \min_{\boldsymbol{\beta}} \|\boldsymbol{Y}-X\boldsymbol{\beta}\|^2$$

a hiba négyzetösszeg feltételes, illetve feltétel nélküli minimuma. Legyen $\operatorname{rang} X = r$, $\operatorname{rang}(F-F_0) = q$. *(Itt $F-F_0$ az F_0 altérnek az F altérre vett ortogonális komplementerét jelöli.)*

Ekkor

$$\frac{S_B^2 - S_e^2}{q} : \frac{S_e}{n-r} \sim F_{q,n-r}$$

akkor és csak akkor, ha H_0 igaz.

BIZONYÍTÁS. Tekintsük az

$$\boldsymbol{Y}-P_{F_0}\boldsymbol{Y} = (\boldsymbol{Y}-P_F\boldsymbol{Y}) + (P_F\boldsymbol{Y}-P_{F_0}\boldsymbol{Y})$$

felbontást. Mivel $F_0 \subseteq F$, így $P_F\boldsymbol{Y}-P_{F_0}\boldsymbol{Y}$ benne van F-ben. $\boldsymbol{Y}-P_F\boldsymbol{Y}$ viszont merőleges F-re. Tehát a fenti felbontás ortogonális felbontás. Ezért

$$\|\boldsymbol{Y}-P_{F_0}\boldsymbol{Y}\|^2 = \|\boldsymbol{Y}-P_F\boldsymbol{Y}\|^2 + \|P_F\boldsymbol{Y}-P_{F_0}\boldsymbol{Y}\|^2.$$

Másrészt

$$S_B^2 = \min_{\boldsymbol{\beta}\,:\,B\boldsymbol{\beta}=\boldsymbol{0}} \|\boldsymbol{Y}-X\boldsymbol{\beta}\|^2 = \min_{\boldsymbol{\beta}\,:\,X\boldsymbol{\beta}\in F_0} \|\boldsymbol{Y}-X\boldsymbol{\beta}\|^2 = \|\boldsymbol{Y}-P_{F_0}\boldsymbol{Y}\|^2$$

és

$$S_e^2 = \min_{\boldsymbol{\beta}} \|\boldsymbol{Y} - X\boldsymbol{\beta}\|^2 = \|\boldsymbol{Y} - P_F\boldsymbol{Y}\|^2 = \|(I - P_F)\boldsymbol{Y}\|^2.$$

Így

$$S_B^2 - S_e^2 = \|P_F\boldsymbol{Y} - P_{F_0}\boldsymbol{Y}\|^2 = \|(P_F - P_{F_0})\boldsymbol{Y}\|^2.$$

Tehát az $S_B^2 - S_e^2$-ben szereplő lineáris kifejezés $(P_F - P_{F_0})\boldsymbol{Y}$, az S_e^2-ben szereplő pedig $(I - P_F)\boldsymbol{Y}$. Ezek együttesen normális eloszlásúak. Korrelálatlanok, mert $(I - P_F)(P_F - P_{F_0}) = 0$. Ezért függetlenek.

$$\mathbb{E}(I - P_F)\boldsymbol{Y} = (I - P_F)X\boldsymbol{\beta} = \boldsymbol{0}$$

mindig teljesül. Viszont

$$\mathbb{E}(P_F - P_{F_0})\boldsymbol{Y} = P_F X\boldsymbol{\beta} - P_{F_0}X\boldsymbol{\beta} = X\boldsymbol{\beta} - P_{F_0}X\boldsymbol{\beta} = \boldsymbol{0} \iff B\boldsymbol{\beta} = \boldsymbol{0},$$

mivel ez utóbbi $X\boldsymbol{\beta} \in F_0$-lal ekvivalens.

Tehát $\frac{1}{\sigma^2}S_e^2 \sim \chi^2_{n-r}$ (ahol r az X rangja), hiszen $I - P_F$ $(n-r)$-dimenziós altérbe vetít.

Ha $H_0 : B\boldsymbol{\beta} = \boldsymbol{0}$ teljesül, akkor $\frac{1}{\sigma^2}(S_B^2 - S_e^2) \sim \chi^2_q$, hiszen $P_F - P_{F_0}$ q-dimenziós altérbe vetít. □

2.3. Az egyszeres osztályozás mint speciális lineáris modell

A lineáris modell alapján elvégezzük az egyszeres osztályozás elemzését.

2.1. példa. Legyen $Y_{ij} = a_i + \varepsilon_{ij}$, $j = 1, \dots, n_i$, $i = 1, \dots, p$, ahol $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ és a ε_{ij} változók függetlenek. Ekkor

$$\boldsymbol{Y} = X\boldsymbol{a} + \boldsymbol{\varepsilon}$$

a modell, ahol

$$\boldsymbol{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{pn_p} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & & & 1 \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{pn_p} \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}.$$

Itt $\boldsymbol{Y}$ és $\boldsymbol{\varepsilon}$ n-dimenziós vektorok, X pedig $n \times p$ méretű mátrix, ahol $n = n_1 + \dots + n_p$.

A

$$H : a_1 = a_2 = \dots = a_p$$

nullhipotézis akkor és csak akkor teljesül, ha $B\boldsymbol{a} = \boldsymbol{0}$, ahol

$$B = \begin{pmatrix} 1 & & 0 & -1 \\ & \ddots & & \vdots \\ 0 & & 1 & -1 \end{pmatrix}$$

$(p-1)\times p$ típusú mátrix. B sorai becsülhetőek, hiszen $B = B^*X$, ahol

$$B^* = \begin{pmatrix} 1 & 0 & \dots & & \dots & 0 & -1 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 & -1 \\ \vdots & & & & & & & \end{pmatrix}.$$

$$S_e^2 = \min_{\boldsymbol{a}} \|Y - X\boldsymbol{a}\|^2 = \min_{\boldsymbol{a}} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - a_i)^2.$$

Ennek minimumhelye (deriválással meghatározva) $\widehat{a}_i = \overline{Y}_{i\cdot}$, azaz az i-edik szinten vett átlag. Így $S_e^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2$. Ez éppen a Q_2 hiba négyzetösszeg. Ennek szabadsági foka a 2.2. tétel alapján $n-p$, mivel X rangja p.

$$S_B^2 = \min_{\boldsymbol{a}:\, B\boldsymbol{a}=\boldsymbol{0}} \sum_i \sum_j (Y_{ij} - a_i)^2.$$

Multiplikátorokkal dolgozunk. $\lambda_1, \dots, \lambda_p$ a multiplikátorok. A Lagrange-függvény:

$$\sum_i \sum_j (Y_{ij} - a_i)^2 + \sum_{i=1}^{p-1} \lambda_i (a_i - a_p) \qquad \text{és} \quad a_i = a_p,\ i = 1, \dots, p-1.$$

a_i szerint deriválva a fenti függvényt:

$$-2 \sum_j (Y_{ij} - a_i) + \lambda_i = 0, \qquad i = 1, \dots, p-1,$$

és

$$-2 \sum_j (Y_{pj} - a_p) - \sum_{i=1}^{p-1} \lambda_i = 0.$$

Ezeket összeadva:

$$\sum_{i=1}^{p-1} \sum_{j=1}^{n_i} (Y_{ij} - a_i) + \sum_{j=1}^{n_p} (Y_{pj} - a_p) = 0.$$

Felhasználva az $a_i = a_p$ $(i = 1, \dots, p-1)$ feltételt: $a_p = \frac{1}{n} \sum_i \sum_j Y_{ij} = \overline{Y}_{\cdot\cdot}$, azaz minden a_i becslése most az $\overline{Y}_{\cdot\cdot}$ teljes átlag. Tehát

$$S_B^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{\cdot\cdot})^2 = Q$$

teljes négyzetösszeg.

$$S_B^2 - S_e^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{\cdot\cdot})^2 - \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2 =$$
$$= Q - Q_2 = Q_1 = \sum_i \sum_j (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2.$$

Azaz $S_B^2 - S_e^2 = Q_1$, a szintek közötti eltérést mérő négyzetösszeg.

$B^*{}_{|F}$ nulltere az azonos koordinátájú vektorokból áll, ami egydimenziós. F az X oszlopai által generált altér. Ezért F p-dimenziós. Így $\operatorname{rang}(F - F_0) = p-1$.

Tehát az egyszeres osztályozás elemezhető a lineáris modell segítségével. A 2.3. tétel alapján adódó F-próba ugyanaz lesz, mint a korábban a Fischer-Cochran-tétel alapján kapott F-próba. □

III. fejezet
A maximum-likelihood módszer

1. A Rao-Cramér-féle egyenlőtlenség

1.1. A likelihood- és a loglikelihood-függvény. A regularitási feltételek

Felsoroljuk azokat a regularitási feltételeket, amelyek esetén a Rao-Cramér-egyenlőtlenséget bizonyítani fogjuk.

Legyen a Θ paramétertér nyílt halmaz $\mathbb{R}^k$-ban. Legyen $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)^\top \in \Theta$.
Legyen X minta (általános értelemben).
Legyen $f(x, \boldsymbol{\vartheta})$ a minta sűrűségfüggvénye (általános értelemben, azaz egy ν dominált mértékre nézve).

Tegyük fel, hogy az $f(x, \boldsymbol{\vartheta})$ függvény $\boldsymbol{\vartheta}$ szerint kétszer differenciálható, és az

$$\int f(x, \boldsymbol{\vartheta}) d\nu(x) = 1$$

kifejezésben a $\boldsymbol{\vartheta}$ szerinti első és második deriválás és az integrálás sorrendje fölcserélhető.

Ekkor

$$\int \frac{\partial f(x, \boldsymbol{\vartheta})}{\partial \vartheta_i} d\nu(x) = 0, \qquad i = 1, \dots, k, \tag{1.1}$$

és

$$\int \frac{\partial^2 f(x, \boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} d\nu(x) = 0, \qquad i, j = 1, \dots, k. \tag{1.2}$$

Tegyük fel, hogy

$$\mathbb{E}\left(\frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_i} \right)^2 < \infty, \qquad i = 1, \dots, k. \tag{1.3}$$

Ekkor

$$I_{ij}(\boldsymbol{\vartheta}) = \mathbb{E}\left[\frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_i}, \frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_j} \right] \tag{1.4}$$

véges minden $i, j = 1, \dots, k$ esetén.

1.1. definíció. Az

$$I(\boldsymbol{\vartheta}) = \Big(I_{ij}(\boldsymbol{\vartheta}) \Big)_{i,j=1}^{k}$$

mátrixot **Fisher-féle információs mátrixnak** nevezzük.

Az $f(x, \boldsymbol{\vartheta})$ függvényt **likelihood-függvénynek** hívjuk, az $\ln f(x, \boldsymbol{\vartheta})$ függvényt pedig **loglikelihood-függvénynek**. A

$$\frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \left(\frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_1}, \dots, \frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_k} \right)^\top \tag{1.5}$$

valószínűségi vektorváltozó az ún. **score vektor**. □

A fenti feltételek esetén a score vektor várható értéke $\mathbf{0}$. Ugyanis

$$\mathbb{E}\frac{\partial \log f(X,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \int \frac{1}{f(x,\boldsymbol{\vartheta})}\frac{\partial f(x,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} f(x,\boldsymbol{\vartheta}) d\nu(x) = \mathbf{0} \tag{1.6}$$

az (1.1) feltétel miatt. (Ne tévesszük szem elöl, hogy a várható érték $\mathbb{E}$ jele után az X véletlen mennyiség szerepel az aktuális függvényben, amikor viszont az $\int$ integrálra térünk, akkor pedig az x változó szerepel ugyanabban a függvényben. Természetesen az utóbbi esetben az f sűrűségfüggvénnyel szoroznunk kell, mielőtt a ν mérték szerint integrálnánk.)

1.1. megjegyzés. Az előzőből azonnal adódik, hogy az $I(\boldsymbol{\vartheta})$ Fisher-féle információs mátrix éppen a score vektor szórástrixa:

$$I(\boldsymbol{\vartheta}) = \operatorname{var}\left(\frac{\partial \log f(X,\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right). \tag{1.7}$$

Másrészt a Fisher-féle információs mátrix a score vektor segítségével előáll az

$$I_{ij}(\boldsymbol{\vartheta}) = -\mathbb{E}\left[\frac{\partial^2 \log f(X,\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right] \tag{1.8}$$

szerint is. Valóban, (1.2) felhasználásával

$$\mathbb{E}\frac{\partial^2 \log f(X,\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} = \mathbb{E}\frac{\partial}{\partial \vartheta_j}\left(\frac{1}{f},\frac{\partial f}{\partial \vartheta_i}\right) = \mathbb{E}\left[\frac{1}{f}\frac{\partial^2 f}{\partial \vartheta_i \partial \vartheta_j} - \frac{1}{f^2}\frac{\partial f}{\partial \vartheta_i}\frac{\partial f}{\partial \vartheta_j}\right] =$$

$$= \int \frac{\partial^2 f(x,\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} d\nu(x) - \mathbb{E}\frac{1}{f}\frac{\partial f}{\partial \vartheta_i}\cdot\frac{1}{f}\frac{\partial f}{\partial \vartheta_j} = 0 - I_{ij}(\boldsymbol{\vartheta}). \qquad \square$$

Folytatjuk a Rao-Cramér-egyenlőtlenség feltételeinek felsorolását.

Feltesszük, hogy $I(\boldsymbol{\vartheta})$ invertálható.
Legyen $\boldsymbol{g} : \Theta \to \mathbb{R}^r$ differenciálható függvény.
Legyen a $\boldsymbol{T}$ r-dimenziós statisztika a $\boldsymbol{g}(\boldsymbol{\vartheta})$ torzítatlan becslése:

$$\int \boldsymbol{T}(x) f(x,\boldsymbol{\vartheta}) d\nu(x) = \boldsymbol{g}(\boldsymbol{\vartheta}). \tag{1.9}$$

Tegyük fel, hogy ebben az egyenletben a $\boldsymbol{\vartheta}$ szerinti differenciálás és az integrálás sorrendje felcserélhető:

$$\int \boldsymbol{T}(x)\frac{\partial f(x,\boldsymbol{\vartheta})}{\partial \vartheta_i} d\nu(x) = \frac{\partial \boldsymbol{g}(\boldsymbol{\vartheta})}{\partial \vartheta_i}, \qquad i = 1,\dots,k. \tag{1.10}$$

1.2. A Rao-Cramér-féle egyenlőtlenség

A Rao-Cramér-féle egyenlőtlenség azt állítja, hogy – a regularitási feltételek esetén – a torzítatlan becslés szórása nem lehet akármilyen kicsi. Egy (abszolút) alsó határ adható a Fisher-féle információ segítségével.

1.1. tétel (Rao-Cramér-egyenlőtlenség). *A fenti regularitási feltételek esetén*

$$\boxed{\operatorname{var}(\boldsymbol{T}) \geq G I^{-1}(\boldsymbol{\vartheta}) G^{\top}}, \tag{1.11}$$

ahol $G = \left(\frac{\partial g(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)$.

BIZONYÍTÁS. Tekintsük a

$$\left(T_1(X), \dots, T_r(X),\ \frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_1}, \dots, \frac{\partial \log f(X, \boldsymbol{\vartheta})}{\partial \vartheta_k}\right) \tag{1.12}$$

$(r+k)$ dimenziós statisztika transzponáltját. Ennek szórásmátrixa az előzőek szerint

$$\begin{pmatrix} V & G \\ G^\top & I(\boldsymbol{\vartheta}) \end{pmatrix}, \tag{1.13}$$

ahol $\boldsymbol{T}$ szórásmátrixa V, $G = \left(\frac{\partial g(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)$ pedig $r \times k$ típusú mátrix. Mivel az (1.13) szórásmátrix, így determinánsa nem negatív.

Ezért

$$0 \le \det \begin{pmatrix} E & -GI^{-1}(\boldsymbol{\vartheta}) \\ 0 & I^{-1}(\boldsymbol{\vartheta}) \end{pmatrix} \quad \det \begin{pmatrix} V & G \\ G^\top & I(\boldsymbol{\vartheta}) \end{pmatrix} =$$

$$= \det \begin{pmatrix} V - GI^{-1}(\boldsymbol{\vartheta})G^\top & 0 \\ I^{-1}(\boldsymbol{\vartheta})G^\top & E \end{pmatrix} = \det(V - GI^{-1}(\boldsymbol{\vartheta})G^\top),$$

ahol E és 0 alkalmas méretű egység-, illetve nullmátrixot jelöl.

Ez a gondolatmenet igaz akkor is, ha csupán a $\boldsymbol{T}$ vektor néhány koordinátáját tekintjük. Ezért a $(V - GI^{-1}(\boldsymbol{\vartheta})G^\top)$ mátrix minden, a főátlóra szimmetrikusan elhelyezkedő részmátrixának nem negatív a determinánsa. Tehát $(V - GI^{-1}(\boldsymbol{\vartheta})G^\top)$ pozitív szemidefinit. Ezzel beláttuk a tételt. □

1.1. feladat. Legyen $\xi_1, \dots, \xi_n$ minta $\mathcal{N}(\mu, \sigma^2)$-ből. A paraméter kétdimenziós: $\boldsymbol{\vartheta} = (\mu, \sigma^2)^\top$. Lássuk be az alábbiakat.

(a) A Fisher-féle információs mátrix:

$$I(\boldsymbol{\vartheta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \quad I^{-1}(\boldsymbol{\vartheta}) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}. \tag{1.14}$$

(b) $\bar{\xi}$ és s_n^{*2} torzítatlan becslés (μ, σ^2)-re. A becslés szórásmátrixa:

$$\mathbb{D}^2 \begin{pmatrix} \bar{\xi} \\ s_n^{*2} \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix}. \tag{1.15}$$

(c) Bár a Rao-Cramér-féle alsó határ nem éretik el, s_n^{*2} mégis a legkisebb szórású torzítatlan becslés σ^2-re.

MEGOLDÁS. (c) Legyen t torzítatlan becslés σ^2-re: $\mathbb{E}t = \sigma^2$. Legyen $u = t - s_n^{*2}$, ahonnan $t = u - s_n^{*2}$.

Viszont t és s_n^{*2} várható értéke σ^2, így u várható értéke 0. De a minta együttes sűrűségfüggvénye alapján:

$$0 = \mathbb{E}u = \int u(\boldsymbol{x}) \cdot f(\boldsymbol{x}) d\boldsymbol{x} = \int u(\boldsymbol{x}) \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}} d\boldsymbol{x}\,, \tag{1.16}$$

ahol $\boldsymbol{x} = (x_1 \dots, x_n)^\top$. μ szerint deriválva kétszer (közben (1.16)-ot még egyszer használva):

$$\int \underbrace{\Big(\sum(x_i-\mu)\Big)^2}_{(n(\bar{x}-\mu))^2} u(\boldsymbol{x}) \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} d\boldsymbol{x} = 0. \tag{1.17}$$

σ^2 szerint deriválva (1.16)-ot (közben (1.16)-ot még egyszer használva):

$$\int \sum(x_i-\mu)^2 u(\boldsymbol{x}) \frac{1}{(2\pi)^{n/2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} d\boldsymbol{x} = 0. \tag{1.18}$$

Viszont a Steiner-formula szerint $(n-1)s_n^{*2} = \sum(x_i-\mu)^2 - n(\bar{x}-\mu)^2$, így az előző két egyenlet miatt u és s_n^{*2} korrelálatlan. Ezért

$$\mathbb{D}^2(t) = \mathbb{D}^2(u) + \mathbb{D}^2(s_n^{*2}) \geq \mathbb{D}^2(s_n^{*2}). \qquad \square$$

2. A maximum-likelihood becslés

2.1. A maximum-likelihood becslés definíciója

2.1. definíció. Legyen X minta (általános értelemben), x a minta realizáció, Θ a paramétertér. Legyen ν mérték a mintatéren, tegyük fel, hogy a minta eloszlása abszolút folytonos ν-re nézve, jelölje $\ell(x,\boldsymbol{\vartheta})$ a sűrűségfüggvényt.
$\ell(x,\boldsymbol{\vartheta})$-t – mint $\boldsymbol{\vartheta}\in\Theta$ függvényét tekintve – **likelihood-függvénynek** nevezzük. □

2.2. definíció. A $\boldsymbol{\vartheta}$ paraméter **maximum-likelihood becslésének** azt a $\widehat{\boldsymbol{\vartheta}}=\widehat{\boldsymbol{\vartheta}}(x)$ statisztikát nevezzük, melyre

$$\ell(x,\widehat{\boldsymbol{\vartheta}})=\sup_{\boldsymbol{\vartheta}\in\Theta}\ell(x,\boldsymbol{\vartheta}). \tag{2.1}$$

□

2.1. megjegyzés. Ha létezik $T(x)$ elégséges statisztika, akkor a maximum-likelihood becslés ennek függvénye. Valóban, a Neyman-Fisher-féle faktorizációs tétel szerint

$$\ell(x,\boldsymbol{\vartheta})=h(x)g(T(x),\boldsymbol{\vartheta}),$$

így a maximum hely csak $T(x)$-től függ.

A fenti állítás megfordítása nem igaz, a maximum-likelihood becslés nem mindig elégséges statisztika. □

2.3. definíció. A likelihood-függvény logaritmusát **loglikelihood-függvénynek** nevezzük:

$$L(x,\boldsymbol{\vartheta})=\log\ell(x,\boldsymbol{\vartheta}). \tag{2.2}$$

□

Ha $\Theta\subseteq\mathbb{R}^k$ nyílt halmaz és $\ell(x,\boldsymbol{\vartheta})$ $\boldsymbol{\vartheta}$ szerint differenciálható, akkor $L(x,\boldsymbol{\vartheta})$ maximum helyén a $\boldsymbol{\vartheta}$ szerinti parciális deriváltak eltünnek.

2.4. definíció. A

$$\frac{\partial L(x,\boldsymbol{\vartheta})}{\partial\vartheta_i}=0,\qquad i=1,\dots,k, \tag{2.3}$$

egyenleteket **likelihood-egyenleteknek** nevezzük. □

2.2. megjegyzés. Ha $\boldsymbol{x}$ független valószínüségi változókból álló vektor, akkor a loglikelihood-függvény:

$$L_n=L(\boldsymbol{x},\boldsymbol{\vartheta})=\sum_{i=1}^{n}\log f_i(x_i,\boldsymbol{\vartheta}), \tag{2.4}$$

ahol $f_i(x_i,\boldsymbol{\vartheta})$ az $\boldsymbol{x}$ vektor i-edik koordinátájának a sűrűségfüggvénye. □

2.2. A maximum-likelihood módszer korlátai

2.3. megjegyzés. Lehetséges, hogy a likelihood-függvény nem korlátos felülről, azaz nem létezik maximum-likelihood becslés. □

2.4. megjegyzés. Lehetséges, hogy a likelihood-egyenleteknek van olyan gyöke, mely nem konzisztens becslése a paraméternek. Ezt az alábbi példa mutatja. □

2.1. példa. Legyen $X_1, \dots, X_n$ minta $\mathcal{N}(\vartheta, c^2\vartheta^2)$ eloszlásból, ahol $c \neq 0$ ismert, ϑ ismeretlen paraméter ($\vartheta \neq 0$). Ekkor a loglikelihood függvény:

$$L_n(\vartheta) = n \log \frac{1}{\sqrt{2\pi}(c\vartheta)} - \sum_{i=1}^{n} \frac{(x_i - \vartheta)^2}{2c^2\vartheta^2}.$$

$$\frac{\partial L_n(\vartheta)}{\partial \vartheta} = -n\vartheta^{-1} + \frac{1}{2c^2} \sum_{i=1}^{n} \left[2(x_i - \vartheta)\vartheta^{-2} + (x_i - \vartheta)^2 2\vartheta^{-3} \right].$$

$$\frac{\partial L_n(\vartheta)}{\partial \vartheta} = 0$$

alapján

$$nc^2\vartheta^2 + \vartheta \sum\nolimits_{i=1}^{n} x_i - \sum\nolimits_{i=1}^{n} x_i^2 = 0.$$

Innen

$$\widehat{\vartheta}_{12} = \frac{-\sum_{i=1}^{n} X_i \pm \sqrt{\left(\sum_{i=1}^{n} X_i\right)^2 + 4nc^2 \sum_{i=1}^{n} X_i^2}}{2nc^2}$$

a megoldás.

Ezen statisztikák határértékét $n \to \infty$ estén a nagy számok törvénye alapján határozzuk meg:

$$\frac{\sum_{i=1}^{n} X_i}{n} \to \vartheta \quad \text{és} \quad \frac{\sum_{i=1}^{n} X_i^2}{n} \to \vartheta^2(c^2 + 1).$$

Így $\widehat{\vartheta}_1 \to \vartheta$ és $\widehat{\vartheta}_2 \to -\vartheta\left(1 + \frac{1}{c^2}\right)$.

Tehát ϑ_1 a likelihood-egyenletnek konzinsztens gyöke, ϑ_2 pedig nem konzisztens gyöke. □

2.5. megjegyzés. Lehetséges, hogy a maximum-likelihood becslés nem konzisztens. Erre utal az alábbi példa. □

2.2. példa. Legyenek $\begin{pmatrix} Y_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ független, kétdimenziós véletlen vektorok,

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right), \qquad i = 1, \dots, n.$$

Ekkor a likelihood-függvény:

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i - \mu_i)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}}.$$

A loglikelihood-függvény:

$$L = -2n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\mu_i)^2+(y_i-\mu_i)^2\right].$$

$$\frac{\partial L}{\partial \mu_i} = \frac{1}{\sigma^2}\left[(x_i-\mu_i)+(y_i-\mu_i)\right],$$

ahonnan $\widehat{\mu}_i = \frac{X_i+Y_i}{2}$.

$$\frac{\partial L}{2\sigma^2} = -n(\sigma^2)^{-1} + (\sigma^2)^{-2}\frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\mu_i)^2+(y_i-\mu_i)^2\right].$$

Innen

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n}\left[(X_i-\widehat{\mu}_i)^2+(Y_i-\widehat{\mu}_i)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-Y_i}{2}\right)^2.$$

Itt $\mathbb{E}\left(\frac{X_i-Y_i}{2}\right)^2 = \frac{\sigma^2}{2}$, ahonnan a nagy számok törvénye értelmében $\widehat{\sigma} \to \frac{\sigma^2}{2}$ $(n\to\infty)$, azaz a becslés nem konzisztens.
Megjegyezzük, hogy a példában a paraméterek száma végtelenhez tart, ha $n\to\infty$. □

2.6. megjegyzés. Lehetséges, hogy a likelihood-egyenlet gyökei nem a likelihood-függvény maximumhelyét szolgáltatják. □

2.3. Egy konzisztencia tétel

Az alábbiakban feltesszük, hogy a mintaelemek függetlenek és azonos eloszlásúak. Tegyük fel, hogy különböző paraméter értékekhez különböző eloszlások tartoznak. Jelölje P_ϑ, illetve $\mathbb{E}_\vartheta$ a ϑ paraméternek megfelelő eloszlást, illetve várható értéket. Jelölje $f(x,\vartheta)$ egy mintaelem sűrűségfüggvényét. Jelölje ϑ^* a paraméter valódi értékét.

2.1. tétel. *Legyen $\Theta\subseteq\mathbb{R}$ és legyen $f(x,\vartheta)$ deriválható ϑ szerint egy olyan intervallumon, mely ϑ^*-ot tartalmazza. Ekkor a likelihood-egyenletnek létezik olyan gyöke, mely $n\to\infty$ esetén ϑ^*-hoz tart (P_{ϑ^*} szerint) 1 valószínűséggel.*

Bizonyítás. Tegyük fel, hogy $\log f(x,\vartheta)$ deriválható a $(\vartheta^*-\delta,\vartheta^*+\delta)$ intervallumon. Ismeretes, hogy $\log t \le t-1$. Itt t helyére $\frac{f(x,\vartheta^*\mp\delta)}{f(x,\vartheta^*)}$-ot helyettesítve:

$$\int f(x,\vartheta^*)\log\frac{f(x,\vartheta^*-\delta)}{f(x,\vartheta^*)}d\nu(x) < 0,$$

$$\int f(x,\vartheta^*)\log\frac{f(x,\vartheta^*+\delta)}{f(x,\vartheta^*)}d\nu(x) < 0.$$

Azaz

$$\mathbb{E}_{\vartheta^*}\log f(X,\vartheta^*-\delta) < \mathbb{E}_{\vartheta^*}\log f(X,\vartheta^*),$$

$$\mathbb{E}_{\vartheta^*}\log f(X,\vartheta^*+\delta) < \mathbb{E}_{\vartheta^*}\log f(X,\vartheta^*).$$

Mivel L_n független, azonos eloszlású valószínűségi változók összege, azaz

$$L_n(X_1,\dots,X_n,\vartheta)=\sum_{i=1}^{n}\log f(X_i,\vartheta),$$

így a nagy számok törvénye alkalmazható. Tehát

$$\lim_{n\to\infty}\frac{1}{n}L_n(X_1,\dots,X_n,\vartheta^*+\delta)<\lim_{n\to\infty}\frac{1}{n}L_n(X_1,\dots,X_n,\vartheta^*),$$

$$\lim_{n\to\infty}\frac{1}{n}L_n(X_1,\dots,X_n,\vartheta^*-\delta)<\lim_{n\to\infty}\frac{1}{n}L_n(X_1,\dots,X_n,\vartheta^*)$$

P_{ϑ^*} szerint 1 valószínűséggel.

Ezért elég nagy n-re L_n értéke a $(\vartheta^*-\delta,\vartheta^*+\delta)$ intervallum végpontjaiban kisebb, mint ϑ^*-ban. Így L_n-nek van a $(\vartheta^*-\delta,\vartheta^*+\delta)$-ban lokális maximuma. Mivel $\delta>0$ tetszőleges, így van 1 valószínűséggel konvergens gyök. □

2.4. A maximum-likelihood becslés aszimptotikája heurisztikus magyarázattal

Legyen $\boldsymbol{\vartheta}\in\Theta\subseteq\mathbb{R}^k$ a paraméter, $\boldsymbol{\vartheta}^*$ pedig ennek a valódi értéke. Az $X_1,\dots,X_n$ független, azonos eloszlású mintaelemek esetén a loglikelihood függvény is független, azonos eloszlású tagok összege:

$$L_n(\boldsymbol{\vartheta})=\sum_{i=1}^{n}\log f(X_i,\boldsymbol{\vartheta}), \tag{2.5}$$

ahol f a közös sűrűségfüggvény. Így a score vektor:

$$\frac{\partial L_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}=\sum_{i=1}^{n}\frac{\partial\log f(X_i,\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}. \tag{2.6}$$

Itt az összeadandók független, azonos eloszlású, $\mathbf{0}$ várható értékű és $I(\boldsymbol{\vartheta})$ szórásmátrixú változók. (Itt $I(\boldsymbol{\vartheta})$ egyetlen megfigyelés esetén a Fisher-féle információs mátrix.) A központi határeloszlás tétel miatt $n\to\infty$ esetén

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}\implies b,\quad\text{ahol}\quad b\sim\mathcal{N}_k(\mathbf{0},I(\boldsymbol{\vartheta})). \tag{2.7}$$

Másrészt a loglikelihood-függvény második deriváltjából álló mátrix:

$$\frac{\partial^2 L_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}=\sum_{i=1}^{n}\frac{\partial^2\log f(X_i,\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}. \tag{2.8}$$

Ez független, $-I(\boldsymbol{\vartheta})$ várható értékű tagokból áll, így a nagy számok törvénye miatt:

$$\frac{1}{n}\frac{\partial^2 L_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}\to -I(\boldsymbol{\vartheta}) \tag{2.9}$$

majdnem biztosan.

Fejtsük Taylor-sorba a loglikelihood-függvényt a $\boldsymbol{\vartheta}^*$ igazi paraméter körül:

$$L_n(\boldsymbol{\vartheta}) \approx L_n(\boldsymbol{\vartheta}^*) + \frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\vartheta}^*)^\top}{\partial \boldsymbol{\vartheta}} \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*) +$$

$$+ \frac{1}{2} \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)^\top \left[\frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} \right] \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*). \tag{2.10}$$

A maradék tagot elhagytuk, az n és a $\sqrt{n}$ szorzókat utólag írtuk be. Az előzőek alapján a központi határeloszlás tétel és a nagy számok törvénye alkalmazható. Ezért (2.10) aszimptotikusan

$$L_n(\boldsymbol{\vartheta}) \approx c + \boldsymbol{b}^\top \boldsymbol{y} - \frac{1}{2} \boldsymbol{y}^\top I(\boldsymbol{\vartheta}^*) \boldsymbol{y}, \tag{2.11}$$

ahol $\boldsymbol{y} = \sqrt{n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)$, $\boldsymbol{b} \sim \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*))$, $c = L_n(\boldsymbol{\vartheta}^*)$.

A (2.11) kifejezés más alakban:

$$L_n(\boldsymbol{\vartheta}) \approx c + \frac{1}{2} \boldsymbol{b}^\top I^{-1}(\boldsymbol{\vartheta}^*) \boldsymbol{b} - \frac{1}{2} \left(I^{\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{y} - I^{-\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{b} \right)^\top \left(I^{\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{y} - I^{-\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{b} \right).$$

Ezen kifejezés maximuma az $\boldsymbol{y} = I^{-1}(\boldsymbol{\vartheta}^*) \boldsymbol{b}$ pontban van, a maximum értéke pedig $c + \frac{1}{2} \boldsymbol{b}^\top I^{-1}(\boldsymbol{\vartheta}^*) \boldsymbol{b}$.

Tehát a $\widehat{\boldsymbol{\vartheta}}_n$ maximum-likelihood becslésre

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \approx I^{-1}(\boldsymbol{\vartheta}^*) \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*)) = \mathcal{N}_k(\mathbf{0}, I^{-1}(\boldsymbol{\vartheta}^*)), \tag{2.12}$$

azaz $\widehat{\boldsymbol{\vartheta}}_n$ **aszimptotikusan normális eloszlású**. Innen

$$\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^* \approx \mathcal{N}_k(\mathbf{0}, (nI(\boldsymbol{\vartheta}^*))^{-1}). \tag{2.13}$$

Ebből látszik, hogy $\widehat{\boldsymbol{\vartheta}}_n$ szórása 0-hoz tart, várható értéke viszont $\boldsymbol{\vartheta}^*$-hoz, tehát **konzisztens becslés**. Másrészt $\widehat{\boldsymbol{\vartheta}}_n$ **aszimptotikusan legkisebb szórású**, hisz a Rao-Cramér-egyenlőtlenség miatt a (torzítatlan) becslések szórásának alsó határa $(nI(\boldsymbol{\vartheta}^*))^{-1}$. (Itt kihasználtuk, hogy n db független, azonos eloszlású megfigyeléshez tartozó Fisher-információ az 1 db megfigyeléshez tartozónak n-szerese.)

A loglikelihood-függvény aszimptotikája:

$$\max_{\boldsymbol{\vartheta}} 2 \left[L_n(\boldsymbol{\vartheta}) - L_n(\boldsymbol{\vartheta}^*) \right] \approx \boldsymbol{b}^\top I^{-1}(\boldsymbol{\vartheta}^*) \boldsymbol{b} = \left[I^{-\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{b} \right]^\top \left[I^{-\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{b} \right]. \tag{2.14}$$

Mivel $I^{-\frac{1}{2}}(\boldsymbol{\vartheta}^*) \boldsymbol{b} \sim \mathcal{N}_k(\mathbf{0}, E)$, tehát a fenti kifejezés határeloszlása χ^2_k. Ezen fog alapulni a **likelihood-hányados próba** határeloszlása.

2.7. megjegyzés. A likelihood egyenletből is adódik Taylor-sorfejtéssel az aszimptotika:

$$\mathbf{0} = \frac{1}{\sqrt{n}} \frac{\partial L_n(\widehat{\boldsymbol{\vartheta}}_n)}{\partial \boldsymbol{\vartheta}} \approx \underbrace{\frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta}}}_{\Downarrow} + \underbrace{\frac{1}{n} \frac{\partial^2 L_n(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}}}_{\downarrow} \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*). \tag{2.15}$$

$$\mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*)) \qquad -I(\boldsymbol{\vartheta}^*)$$

Tehát $n \to \infty$ esetén $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \approx \mathcal{N}_k(\mathbf{0}, I^{-1}(\boldsymbol{\vartheta}^*))$. □

2.5. A maximum-likelihood becslés aszimptotikája

Bebizonyítjuk, hogy független, X-szel azonos eloszlású, $X_1, \dots, X_n$ mintaelemek esetén a likelihood-egyenletnek van olyan $\widehat{\boldsymbol{\vartheta}}_n$ gyöke, mely erősen konzisztens, aszimptotikusan normális eloszlású és aszimptotikusan legkisebb szórású. Legyen $f(x, \boldsymbol{\vartheta})$ az X sűrűségfüggvénye (általános értelemben, tehát pl. f jelölhet diszkrét eloszlást is).

2.2. tétel. *Tegyük fel, hogy*

I. $\Theta \subseteq \mathbb{R}^k$;

II. *A $\boldsymbol{\vartheta}^*$ valódi paraméter egy ω környezetében léteznek az $f(x, \boldsymbol{\vartheta})$ sűrűségfüggvénynek $\boldsymbol{\vartheta}$ szerinti harmadik parciális deriváltjai;*

III. $\mathbb{E}_{\boldsymbol{\vartheta}} \left(\dfrac{\partial}{\partial \vartheta_j} \log f(X, \boldsymbol{\vartheta}) \right)^2 < \infty, \quad j = 1, \dots, k;$

IV. $\mathbb{E}_{\boldsymbol{\vartheta}} \dfrac{\partial}{\partial \vartheta_j} \log f(X, \boldsymbol{\vartheta}) = 0, \quad j = 1, \dots, k;$

V. *az $I(\boldsymbol{\vartheta}) = (I_{ij}(\boldsymbol{\vartheta}))_{i,j=1}^k$ Fisher-féle információs mátrix pozitív definit $\boldsymbol{\vartheta} \in \omega$ esetén, ahol*

$$I_{ij}(\boldsymbol{\vartheta}) = \mathbb{E}_{\boldsymbol{\vartheta}} \left(\frac{\partial}{\partial \vartheta_i} \log f(X, \boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_j} \log f(X, \boldsymbol{\vartheta}) \right);$$

VI. *léteznek olyan $M_{ij\ell}$ függvények, hogy*

$$\left| \frac{\partial^3}{\partial \vartheta_i \partial \vartheta_j \partial \vartheta_\ell} \log f(x, \vartheta) \right| \le M_{ij\ell}(x),$$

és $m_{ij\ell} = \mathbb{E}_{\boldsymbol{\vartheta}^*} M_{ij\ell}(X) < \infty, \ i, j, \ell = 1, \dots, k.$

Ekkor a $\frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}) = 0$ likelihood-egyenletnek létezik olyan $\widehat{\boldsymbol{\vartheta}}_n$ gyöke, melyre $n \to \infty$ esetén:

(a) $\boxed{\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}^* \quad 1 \text{ valószínűséggel } P_{\boldsymbol{\vartheta}^*} \text{ szerint}}$;

(b) $\boxed{\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \text{ aszimptotikusan } \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*)^{-1}) \text{ eloszlású}}$.

Bizonyítás. (a) Fejtsük $L_n(\boldsymbol{\vartheta})$-t Taylor-sorba $\boldsymbol{\vartheta}^*$ körül.

$$\frac{1}{n} L_n(\boldsymbol{\vartheta}) - \frac{1}{n} L_n(\boldsymbol{\vartheta}^*) = \frac{1}{n} \boldsymbol{A}^\top (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*) + \frac{1}{2n} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)^\top B (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*) +$$

$$+ \frac{1}{6n} \sum_i \sum_j \sum_\ell (\vartheta_i - \vartheta_i^*)(\vartheta_j - \vartheta_j^*)(\vartheta_\ell - \vartheta_\ell^*) \sum_{s=1}^n \gamma_{ij\ell}(X_s) M_{ij\ell}(X_s) =$$

$$= S_1 + S_2 + S_3,$$

ahol

$$\boldsymbol{A} = \frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta})_{|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}} = \left(\frac{\partial L_n(\boldsymbol{\vartheta}^*)}{\partial \vartheta_1}, \dots, \frac{\partial L_n(\boldsymbol{\vartheta}^*)}{\partial \vartheta_k} \right),$$

$$B = \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta})_{|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}} = \left(\frac{\partial^2 L_n(\boldsymbol{\vartheta}^*)}{\partial \vartheta_i \partial \vartheta_j} \right)_{i,j=1}^{k},$$

$$|\gamma_{ij\ell}(X)| \le 1.$$

Mivel

$$\boldsymbol{A} = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\vartheta}} \log f(X_i, \boldsymbol{\vartheta})_{|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}}$$

és

$$B = \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} \log f(X_i, \boldsymbol{\vartheta})_{|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*}},$$

így a nagy számok törvénye miatt $\frac{1}{n}\boldsymbol{A} \to \mathbf{0}$ és $\frac{1}{n}B \to -I(\boldsymbol{\vartheta}^*)$ 1 valószínűséggel $P_{\boldsymbol{\vartheta}^*}$ szerint. Hasonlóan, $\frac{1}{n}\sum_{s=1}^{n} M_{ij\ell}(X_s) \to m_{ij\ell}$ $P_{\boldsymbol{\vartheta}^*}$ szerint 1 valószínűséggel. Azon esemény, melyre mindhárom konvergencia teljesül, 1 valószínűségű. Az alábbiak ezen esemény egy elemi eseményére vonatkoznak. Jelölje c az $I(\boldsymbol{\vartheta}^*)$ legkisebb sajátértékét. Legyen $\varepsilon{>}0$ és $b{=}\frac{k^3}{3}\sum_i \sum_j \sum_\ell m_{ij\ell}$. Elég nagy n-re $\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*\|{=}\varepsilon$ esetén igaz, hogy $|S_1| < \varepsilon^3$, $S_2 < -\frac{c}{4}\varepsilon^2$, $|S_3| < b\varepsilon^3$.

Így a $\boldsymbol{\vartheta}^*$ középpontú, ε sugarú gömb felületén

$$\frac{1}{n}L_n(\boldsymbol{\vartheta}) - \frac{1}{n}L_n(\boldsymbol{\vartheta}^*) = S_1 + S_2 + S_3 < \varepsilon^3 - \frac{c}{4}\varepsilon^2 + b\varepsilon^3 < 0,$$

ha $\varepsilon < \frac{c}{4(1+b)}$.

Tehát ezen a gömbön belül van lokális maximuma $L_n(\boldsymbol{\vartheta})$-nak. Mivel ε tetszőleges pozitív szám, ez a lokális maximum a likelihod-egyenlet $\boldsymbol{\vartheta}^*$-hoz konvergáló gyöke. Jelölje ezt a gyököt $\widehat{\boldsymbol{\vartheta}}_n$.

(b) Fejtsük Taylor-sorba $\frac{\partial}{\partial \vartheta_j}L_n(\widehat{\boldsymbol{\vartheta}}_n)$-et $\boldsymbol{\vartheta}^*$ körül:

$$\frac{\partial}{\partial \vartheta_j}L_n(\widehat{\boldsymbol{\vartheta}}_n) - \frac{\partial}{\partial \vartheta_j}L_n(\boldsymbol{\vartheta}^*) = \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \vartheta_j}L_n(\boldsymbol{\vartheta}^*)^\top(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) +$$

$$+\frac{1}{2}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top \frac{\partial^3}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta} \partial \vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*), \tag{2.16}$$

ahol $\widetilde{\boldsymbol{\vartheta}}$ a $\boldsymbol{\vartheta}^*$ és a $\widehat{\boldsymbol{\vartheta}}_n$ pontokat összekötő szakasz eleme. Mivel $\widehat{\boldsymbol{\vartheta}}_n$ maximum hely, így $\frac{\partial}{\partial \vartheta_j}L_n(\widehat{\boldsymbol{\vartheta}}_n) = 0$.

Minden j-re:

$$\Bigg[\underbrace{\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*)^\top}_{\downarrow\ -I(\boldsymbol{\vartheta}^*)\ \text{1 valószínűséggel}} + \underbrace{\frac{1}{2n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)^\top\frac{\partial^3}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})}_{\downarrow\ 0\ \text{1 valószínűséggel}}\Bigg]\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*) =$$

$$= \underbrace{-\frac{1}{\sqrt{n}}\frac{\partial}{\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*)}_{\Downarrow\ \mathcal{N}_k(\mathbf{0},I(\boldsymbol{\vartheta}^*))\ \text{eloszlásban}}.$$

A fenti konvergenciák úgy értendők, hogy az egyenletet minden j-re kiírva, a kapott vektorok, ill. mátrixok konvergálnak $n\to\infty$ esetén. Az első konvergencia a nagy számok törvénye miatt áll fenn. A második a $\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\to\mathbf{0}$ miatt, és mert $\frac{1}{n}\frac{\partial^3}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})$ majorálható egy – a nagy számok törvénye miatt – konvergens szorzattal. A harmadik a centrális határeloszlás tétel következménye.

Tehát a fenti egyenletrendszer bal oldalán a [ ]-ben $-I(\boldsymbol{\vartheta}^*)+\mathrm{o}(1)$ áll, ahol $\mathrm{o}(1)$ egy 1 valószínűséggel 0-hoz tartó sorozatot jelöl. Ennek inverze $-I^{-1}(\boldsymbol{\vartheta}^*)+\mathrm{o}(1)$. Ezzel szorozva mindkét oldalt, Szluckij lemmája alapján adódik:

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)\Rightarrow\mathcal{N}_k(\mathbf{0},I^{-1}(\boldsymbol{\vartheta}^*)).\qquad\square$$

2.8. megjegyzés. Az előző tétel feltételei mellett $\widehat{\boldsymbol{\vartheta}}_n$ lokális maximum (1-hez tartó valószínűséggel). $\square$

2.9. megjegyzés.

$$\text{(2.17)}\qquad \text{(a)}\quad \frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}L_n(\boldsymbol{\vartheta}^*)\right]+\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}L_n(\boldsymbol{\vartheta}^*)\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)$$

és

$$\text{(2.18)}\qquad \text{(b)}\quad \frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}L_n(\boldsymbol{\vartheta}^*)\right]-I(\boldsymbol{\vartheta}^*)\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)$$

sztochasztikusan $\mathbf{0}$-hoz tart.

Bizonyítás. (a) (2.16) alapján

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*)+\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*)^\top\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*) =$$

$$= -\frac{1}{2n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)^\top\frac{\partial^3}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*).$$

Mivel $\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^* \to \mathbf{0}$ sztochasztikusan, $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)$ eloszlásban konvergens, tehát sztochasztikusan korlátos, továbbá $\frac{1}{n}\frac{\partial^3}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})$ sztochasztikusan korlátos, így a jobb oldal 0-hoz tart sztochasztikusan.

(b) (2.16) alapján

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*) - I_j(\boldsymbol{\vartheta}^*)^\top\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) =$$

$$= -\left[\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*)^\top + I_j(\boldsymbol{\vartheta}^*)^\top\right]\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) -$$

$$- \frac{1}{2n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top\frac{\partial^3}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\widetilde{\boldsymbol{\vartheta}})\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*).$$

Itt $I_j(\boldsymbol{\vartheta}^*)^\top$ az $I(\boldsymbol{\vartheta}^*)$ mátrix j-edik sorvektora. Az előzőekhez még figyelembe kell venni, hogy a nagy számok törvénye miatt $\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\vartheta_j}L_n(\boldsymbol{\vartheta}^*) \to I_j(\boldsymbol{\vartheta}^*)$. □

3. A likelihood-hányados próba

3.1. A loglikelihood-függvény aszimptotikája

Ahhoz, hogy a likelihood-függvényen alapuló próbák statisztikáinak határeloszlását megkapjuk, szükségünk van az L_n függvény aszimptotikájának ismeretére. Legyen $X_1, \dots, X_n$ (független, azonos eloszlású elemekből álló) minta, $L_n(\boldsymbol{\vartheta})$ a megfelelő loglikelihood-függvény, $\widehat{\boldsymbol{\vartheta}}_n$ pedig a likelihood-egyenlet konzisztens gyöke.

3.1. tétel. *A 2.2. tétel feltételeinek teljesülése esetén, ha* $n \to \infty$,

$$\text{(A)} \quad \Big[L_n(\widehat{\boldsymbol{\vartheta}}_n) - L_n(\boldsymbol{\vartheta}^*)\Big] - \frac{1}{2} n (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top I(\boldsymbol{\vartheta}^*)(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \to 0 \tag{3.1}$$

sztochasztikusan $P_{\boldsymbol{\vartheta}^*}$ *szerint;*

$$\text{(B)} \quad 2\Big[L_n(\widehat{\boldsymbol{\vartheta}}_n) - L_n(\boldsymbol{\vartheta}^*)\Big] \implies \chi^2_k \tag{3.2}$$

eloszlásban $P_{\boldsymbol{\vartheta}^*}$ *szerint.*

Bizonyítás. (A) Fejtsük Taylor-sorba $L_n(\boldsymbol{\vartheta})$-t a $\boldsymbol{\vartheta}^*$ körül:

$$L_n(\widehat{\boldsymbol{\vartheta}}_n) - L_n(\boldsymbol{\vartheta}^*) = \frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}^*)^\top (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) +$$

$$+ \frac{1}{2} (\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}^*)(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) + \gamma \|\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\|^3 \sum_{s=1}^n G(X_s),$$

ahol $G(x) = \sum\limits_{i,j,\ell} M_{ij\ell}(x)$ és $|\gamma| \le \frac{1}{6}$. Ebből

$$\begin{aligned}
L_n(\widehat{\boldsymbol{\vartheta}}_n) - L_n(\boldsymbol{\vartheta}^*) &= \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top I(\boldsymbol{\vartheta}^*) \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) + \\
&+ \left[\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}^*)^\top - \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top I(\boldsymbol{\vartheta}^*)\right] \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) + \\
&+ \frac{1}{2} \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top \left[\frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}^*) + I(\boldsymbol{\vartheta}^*)\right] \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) - \\
&- \frac{1}{2} \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)^\top I(\boldsymbol{\vartheta}^*) \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) + \\
&+ \Big[\gamma \|\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\|\Big] \Big[\sqrt{n}\, \|\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\|\Big] \Big[\sqrt{n}\, \|\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\|\Big] \left[\frac{1}{n} \sum_{s=1}^n G(X_s)\right].
\end{aligned}$$

A 2.9. megjegyzés (b) része miatt a második tag sztochasztikusan 0-hoz tart. A harmadik tag ugyanezen megjegyzés (a)–(b) része, valamint $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*)$ sztochasztikus korlátossága miatt 0-hoz tart sztochasztikusan. Az ötödik tag is 0-hoz tart, mivel ennek az első tényezője 0-hoz tart 1 valószínűséggel, míg a következő három tényezője sztochasztikusan korlátos.

Az első és a negyedik tag összege pedig éppen az (A)-ban szereplő mennyiség.

(B) Mivel

$$\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \implies \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*)^{-1}),$$

így (A)-ból következik (B). □

3.2. Egyszerű nullhipotézis vizsgálata

A

$$H_0 : \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$$

egyszerű nullhipotézis

$$H_1 : \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$$

alternatív hipotézissel szembeni tesztelésére a Neyman-Pearson-féle, az A. Waldtól és a C. R. Raotól származó statisztikákat ismertetjük. Ezek mindegyike a maximum-likelihood módszerrel kapcsolatos. Ezen általános statisztikák alapján lehet speciális modellek esetén konkrét próbákat konstruálni.

3.1. definíció. Legyen $X_1, \dots, X_n$ (független, azonos eloszlású elemekből álló) minta, $\boldsymbol{\vartheta} \in \Theta \subseteq \mathbb{R}^k$ a paraméter, $\ell(X_1, \dots, X_n, \boldsymbol{\vartheta})$ a megfelelő likelihood-függvény, $L_n(\boldsymbol{\vartheta}) = L_n(X_1, \dots, X_n, \boldsymbol{\vartheta})$ a loglikelihood-függvény, $I(\boldsymbol{\vartheta})$ a Fisher-féle információs mátrix, $\widehat{\boldsymbol{\vartheta}}_n$ pedig a likelihood-egyenlet konzisztens gyöke. Vezessük be az alábbi statisztikákat.

Neyman-Pearson: $\displaystyle \Lambda_n = \frac{\ell(X_1, \dots, X_n, \boldsymbol{\vartheta}_0)}{\ell(X_1, \dots, X_n, \widehat{\boldsymbol{\vartheta}}_n)}.$

Wald: $\displaystyle W_n = \sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0)^\top I(\widehat{\boldsymbol{\vartheta}}_n)\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0).$

Rao: $\displaystyle V_n = \frac{1}{n}\left[\frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}_0)\right] [I(\boldsymbol{\vartheta}_0)]^{-1} \left[\frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}_0)\right]^\top.$ □

A Neyman-Pearson-féle statisztikát szokták a

$$\Lambda_n = \frac{\ell(X_1, \dots, X_n, \boldsymbol{\vartheta}_0)}{\sup\limits_{\boldsymbol{\vartheta} \in \Theta} \ell(X_1, \dots, X_n, \boldsymbol{\vartheta})}$$

képlettel is definiálni. Az aszimptotikus eloszlás megfogalmazásához azonban az előző definíció alkalmasabb.

A Rao-féle tesztet szokták score tesztnek is nevezni.

3.2. tétel. *A 2.2. tételbeni feltételek esetén $-2\log\Lambda_n$, W_n, valamint V_n is aszimptotikusan χ^2_k eloszlású, ha H_0 igaz.*

Bizonyítás. Az, hogy $-2\log\Lambda_n$ aszimptotikusan χ^2_k eloszlású, azonnal adódik a 3.1. tétel (B) részéből.

Tudjuk, hogy $\sqrt{n}(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*) \implies \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}^*)^{-1})$. Mivel $\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0$ a H_0 teljesülése esetén 1 valószínűséggel, és $I(\boldsymbol{\vartheta})$ $\boldsymbol{\vartheta}$-ben folytonos, így $I(\widehat{\boldsymbol{\vartheta}}_n) \to I(\boldsymbol{\vartheta}_0)$ a H_0 teljesülése esetén 1 valószínűséggel. Tehát W_n is aszimptotikusan χ^2_k eloszlású.

Mivel $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\vartheta}} L_n(\boldsymbol{\vartheta}_0) \implies \mathcal{N}_k(\mathbf{0}, I(\boldsymbol{\vartheta}_0))$, így $V_n \implies \chi^2_k$. □

IV. fejezet
Nem-paraméteres módszerek

1. Hoeffding-féle U-statisztikák

Ebben a részben a Hoeffding-féle U-statisztikákra bizonyítunk határeloszlás-tételt, amiből azután levezetjük a Wilcoxon-féle rangösszeg statisztika aszimptotikus normalitását.

1.1. Független változókkal való közelítés

Legyenek $Z_1, Z_2, \dots, Z_N$ független valószínűségi változók, T pedig legyen olyan valószínűségi változó, melyre $\mathbb{E}(T^2) < \infty$ és $\mathbb{E}(T) = 0$.

Keresendők olyan $k_1, k_2, \dots, k_N$ (Borel-)mérhető függvények, melyekre

$$\mathbb{E}(T - S)^2 \qquad \text{minimális,}$$

ahol

$$S = \sum\nolimits_{i=1}^{N} k_i(Z_i). \tag{1.1}$$

Figyeljük meg, hogy a feladat és annak alábbi megoldása analógiát mutat egy vektornak egy ortogonális vektorrendszer által generált altérre való vetítésével.

1.1. tétel. *A fenti feladat megoldása*

$$k_i(Z_i) = \mathbb{E}(T \mid Z_i), \qquad i = 1, \dots, N.$$

Bizonyítás. Legyen

$$T^* = \sum\nolimits_{i=1}^{N} \mathbb{E}(T \mid Z_i). \tag{1.2}$$

Elegendő belátni, hogy tetszőleges, (1.1)-ben megadott S-re

$$\mathbb{E}(T - S)^2 = \mathbb{E}(T - T^*)^2 + \mathbb{E}(T^* - S)^2. \tag{1.3}$$

Viszont az alábbi alapján a „kétszeres szorzat"$= 0$, amiből (1.3) már következik.

$$\begin{aligned}
\mathbb{E}(T - T^*)(T^* - S) &= \sum\nolimits_{i=1}^{N} \mathbb{E}\left[(T - T^*)\big(\mathbb{E}(T \mid Z_i) - k_i(Z_i)\big)\right] = \\
&= \sum\nolimits_{i=1}^{N} \mathbb{E}\Big[\mathbb{E}\Big\{(T - T^*)\big(\mathbb{E}(T \mid Z_i) - k_i(Z_i)\big)\Big|Z_i\Big\}\Big] = \\
&= \sum\nolimits_{i=1}^{N} \mathbb{E}\Big[\big(\mathbb{E}(T \mid Z_i) - k_i(Z_i)\big)\mathbb{E}\left\{T - T^* \mid Z_i\right\}\Big] = 0\,.
\end{aligned}$$

Itt kihasználtuk, hogy a Z_i és a Z_j $(i \neq j)$ függetlensége alapján

$$\mathbb{E}\big\{T - T^* \mid Z_i\big\} = \mathbb{E}(T \mid Z_i) - \mathbb{E}(T \mid Z_i) - \sum_{i \neq j} \mathbb{E}\left\{\mathbb{E}(T \mid Z_j) \mid Z_i\right\} = 0. \qquad \square$$

1.2. A Hoeffding-féle U-statisztika

Legyen

az X valószínűségi változó eloszlásfüggvénye F,

az Y valószínűségi változó eloszlásfüggvénye G.

Legyen

$$X_1, \dots, X_m \qquad \text{minta } X\text{-re},$$

$$Y_1, \dots, Y_n \qquad \text{minta } Y\text{-ra},$$

ahol feltesszük a két minta egymástól való függetlenségét is.

1.1. definíció. A **Hoeffding-féle U-statisztika**

$$U = \frac{1}{m \cdot n} \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{n} \varphi(X_i, Y_j), \tag{1.4}$$

ahol φ adott kétváltozós mérhető függvény. □

Az a célunk, hogy a Hoeffding-féle U-statisztika (alkalmas normalizáltjának) aszimptotikus normalitását belássuk. Ennek érdekében — a fenti 1.1. tétel segítségével — független valószínűségi változók összegével közelítjük meg U-t. A független összegre már alkalmazhatjuk a központi határeloszlás-tételt. A pontos eljárás az alábbi.

Legyen $\mathbb{E}\varphi^2(X, Y) < \infty$ és legyen

$$\theta = \mathbb{E}U = \mathbb{E}\varphi(X, Y),$$

$$\psi(x, y) = \varphi(x, y) - \theta.$$

Keresendő

$$mn(U - \mathbb{E}U) = \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{n} \psi(X_i, Y_j)$$

legjobb közelítése

$$S = \sum\nolimits_{i=1}^{m} a_i(X_i) + \sum\nolimits_{j=1}^{n} b_j(Y_j)$$

alakú függvényekkel, ahol az a_i és b_j mérhető függvényeket kell meghatározni.

A feltételes várható érték ismert tulajdonsága, hogy ha X és Y függetlenek, akkor $\mathbb{E}\{f(X, Y) \mid X = x\} = \mathbb{E}f(x, Y)$. Ezért

$$\mathbb{E}\{\psi(X_i, Y_j) \mid X_k = x\} = \begin{cases} \mathbb{E}\psi(x, Y_j), & i = k; \\ \mathbb{E}\psi(X_i, Y_j) = 0, & i \neq k. \end{cases}$$

Vezessük be az alábbi jelöléseket:

$$\psi_{10}(x) = \mathbb{E}\psi(x, Y),$$

$$\psi_{01}(y) = \mathbb{E}\psi(X, y).$$

A fenti megjegyzés és az 1.1. tétel alapján $mn(U - \mathbb{E}U)$ keresett közelítése:

$$\sum_{k=1}^{m} \mathbb{E}\left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} \psi(X_i, Y_j) \Big| X_k \right\} + \sum_{\ell=1}^{n} \mathbb{E}\left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} \psi(X_i, Y_j) \Big| Y_\ell \right\} =$$

$$= \sum_{k=1}^{m} n\psi_{10}(X_k) + \sum_{\ell=1}^{n} m\psi_{01}(Y_\ell). \tag{1.5}$$

A következő tételben használni fogjuk az alábbi jelöléseket:

$$\sigma_{10}^2 = \mathbb{D}^2\psi_{10}(X) = \mathbb{E}\psi_{10}^2(X), \qquad \sigma_{01}^2 = \mathbb{D}^2\psi_{01}(Y) = \mathbb{E}\psi_{01}^2(Y).$$

1.2. tétel. *(A Hoeffding-féle U-statisztika aszimptotikus normalitása.) Tegyük fel, hogy $m \le n$ és*

$$\lim_{\substack{n\to\infty \\ m\to\infty}} m/n = \lambda,$$

ahol λ lehet 0 is. Ekkor $m, n \to \infty$ esetén

$$T = \sqrt{m}(U - \theta) \implies \mathcal{N}(0, \sigma_{10}^2 + \lambda\sigma_{01}^2).$$

(Amennyiben a határeloszlás szórása 0, akkor elfajult, pontosabban a 0-ba koncentrált a határeloszlás.)

Bizonyítás. Belátjuk, hogy $T - T^* \to 0$ L^2-ben, és $T^* \implies \mathcal{N}(\,.\,,\,.\,)$, ahol T^* a tételbeni T-hez (1.2) alapján tartozik. Ezekből — Szluckij ismert lemmája alapján — már következni fog az állítás.

Emlékeztetünk Szluckij lemmájára.

Ha $X_n - Y_n \to 0$ sztochasztikusan, és $Y_n \implies Y$ eloszlásban,
akkor $X_n \implies Y$ eloszlásban.

Kezdjük $T - T^* \to 0$ L^2-ben igazolásával. (1.5)-ből:

$$T^* = \sqrt{m}\left[\frac{1}{m}\sum\nolimits_{i=1}^{m}\psi_{10}(X_i) + \frac{1}{n}\sum\nolimits_{j=1}^{n}\psi_{01}(Y_j)\right]. \tag{1.6}$$

(1.3)-ból $S = 0$ vételével:

$$\mathbb{E}(T - T^*)^2 = \mathbb{E}T^2 - \mathbb{E}T^{*2} = \mathbb{D}^2T - \mathbb{D}^2T^*, \tag{1.7}$$

mivel $\mathbb{E}T = \mathbb{E}T^* = 0$. Kiszámítjuk $\mathbb{D}^2(T)$-t és $\mathbb{D}^2(T^*)$-ot.

$$\begin{aligned}
\mathbb{D}^2T &= m\frac{1}{m^2n^2}\sum\nolimits_{k,\ell}\sum\nolimits_{i,j}\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_k,Y_\ell)\big) = \\
&= \frac{1}{mn^2}\Big[\sum\nolimits_{i,j}\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_i,Y_j)\big) + \\
&\quad + \sum_{\substack{i\,j\,\ell \\ \ell\neq j}}\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_i,Y_\ell)\big) + \sum_{\substack{i\,j\,k \\ i\neq k}}\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_k,Y_j)\big)\Big] = \\
&= \frac{1}{mn^2}\big[mn(\dots) + mn(n-1)(\dots) + nm(m-1)(\dots)\big] \to \\
&\underset{m,n\to\infty}{\to} 0 + \operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_i,Y_\ell)\big) + \lambda\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_k,Y_j)\big) = \\
&= \sigma_{10}^2 + \lambda\sigma_{01}^2.
\end{aligned} \tag{1.8}$$

Ugyanis egyrészt a függetlenség miatt $\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_k,Y_\ell)\big)=0$, ha $i\neq k$ és $j\neq\ell$. Másrészt $j\neq\ell$ esetén

$$\operatorname{cov}\big(\varphi(X_i,Y_j),\varphi(X_i,Y_\ell)\big)=\mathbb{E}\big(\psi(X_i,Y_j)\cdot\psi(X_i,Y_\ell)\big)=$$

$$=\mathbb{E}\Big(\underbrace{\mathbb{E}\big\{\psi(X_i,Y_j)\cdot\psi(X_i,Y_\ell)\mid X_i=x\big\}}_{}{}_{|x=X_i}\Big),$$

valamint

$$\mathbb{E}\psi(x,Y_j)\cdot\psi(x,Y_\ell)=\mathbb{E}\psi(x,Y_j)\cdot\mathbb{E}\psi(x,Y_\ell)=\psi_{10}^2(x)\,.$$

Innen $\sigma_{10}^2=\mathbb{E}\psi_{10}^2(X_i)$ már adódik. Hasonlóan σ_{01}^2 is.

Rátérünk $\mathbb{D}^2T^*$-ra. (1.6) alapján, a függetlenséget kihasználva:

$$\mathbb{D}^2T^*=m\left[\frac{1}{m^2}\sum\nolimits_{i=1}^{m}\sigma_{10}^2+\frac{1}{n^2}\sum\nolimits_{j=1}^{n}\sigma_{01}^2\right]=\sigma_{10}^2+\frac{m}{n}\sigma_{01}^2\to\sigma_{10}^2+\lambda\sigma_{01}^2.$$

Így, (1.7) alapján, $T-T^*\to 0$ L^2-ben.

Térjünk rá $T^*\implies\mathcal{N}(\,.\,,\,.\,)$ igazolására. $n,m\to\infty$ esetén

$$T^*=\underbrace{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\psi_{10}(X_i)}_{\Downarrow}+\sqrt{\frac{m}{n}}\;\underbrace{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_{01}(Y_j)}_{\Downarrow}$$

$$\underbrace{\mathcal{N}(0,\sigma_{10}^2)\qquad\lambda\qquad\mathcal{N}(0,\sigma_{01}^2)}_{\Downarrow}$$

$$\mathcal{N}(0,\sigma_{10}^2+\lambda\sigma_{01}^2)\,,$$

ha $\lambda\neq 0$ és legalább σ_{10}^2, σ_{01}^2 egyike pozitív; illetve ha $\lambda=0$ és σ_{10}^2 pozitív (ellenkező esetben T^* (határ)eloszlása a 0-ba koncentrálódik). A fenti határeloszlás meghatározásához a központi határeloszlás tételt és a két összeg egymástól való függetlenségét használtuk ki (lásd a 3. és a 4. Feladatot). □

1.3. Feladatok

1. Az 1.1. tétel bizonyításának mely lépéseinél használtuk ki a feltételes várható érték alábbi tulajdonságait?

(a) $\mathbb{E}\{\xi\mid\eta\}$ az η-nak (Borel-)mérhető függvénye.
(b) $\mathbb{E}(\mathbb{E}\{\xi\mid\eta\})=\mathbb{E}\xi$.
(c) $\mathbb{E}\{\xi\zeta\mid\eta\}=\zeta\mathbb{E}\{\xi\mid\eta\}$, ha ζ η-mérhető.
(d) $\mathbb{E}\{a_1\xi_1+a_2\xi_2\mid\eta\}=a_1\mathbb{E}\{\xi_1\mid\eta\}+a_2\mathbb{E}\{\xi_2\mid\eta\}$, ahol a_1 és a_2 konstansok.
(e) $\mathbb{E}\{\xi\mid\eta\}=\mathbb{E}\xi$, ha ξ és η függetlenek.

2. Ismeretes, hogy $g(y)=\mathbb{E}\{\xi\mid\eta=y\}$ jelöléssel: $\mathbb{E}\{\xi\mid\eta\}=g(\eta)$; vagy tömörebben: $\mathbb{E}\{\xi\mid\eta\}=\big(\mathbb{E}\{\xi\mid\eta=y\}\big)_{|y=\eta}$. A fenti szakaszban mely helyeken használtuk ezt az összefüggést?

3. Legyen ξ_n és η_n független minden n esetén. Tegyük fel, hogy $\xi_n\Rightarrow\xi$ és $\eta_n\Rightarrow\eta$ eloszlásban. Igazoljuk (a karakterisztikus függvények módszerével), hogy ekkor $\xi_n+\eta_n$

eloszlásban konvergál a ξ és η eloszlásának a konvolúciójához. Hogyan alkalmaztuk ezt az 1.2. tétel bizonyításának a végén?

4. Tegyük fel, hogy $\xi_n \Rightarrow \xi$ eloszlásban, $a_n \to a$. Lássuk be, hogy $a_n\xi_n \Rightarrow a\xi$ eloszlásban.

4. ÚTMUTATÁS. Első módszer. Jelölje φ a karakterisztikus függvényt.
Egyrészt $\varphi_{a_n\xi_n}(t) = \mathbb{E}e^{ita_n\xi_n} = \varphi_{\xi_n}(a_nt)$.
Másrészt $|\varphi_{\xi_n}(a_nt) - \varphi_\xi(a_nt)| \le \varepsilon$, ha $n > n_\varepsilon$, mert $\varphi_{\xi_n}(t) \to \varphi_\xi(t)$ minden korlátos intervallumban egyenletesen (l. Rényi (1973)).
Harmadrészt $\varphi_\xi(a_nt) \to \varphi_\xi(at) = \varphi_{a\xi}(t)$, mert φ folytonos.

Második módszer. Alkalmazzuk a Szluckij-lemmát! Ekkor kihasználjuk, hogy $a\xi_n \Rightarrow a\xi$ eloszlásban és $(a_n\xi_n - a\xi_n) \to 0$ sztochasztikusan. Ez utóbbi belátásához szükséges: a ξ_n sorozat sztochasztikusan korlátos (mert eloszlásban konvergens).

5. Érvényben marad-e a Hoeffding-féle U-statisztika aszimptotikus normalitása $m > n$ esetén?

5. MEGOLDÁS. Az 1.2. tétel alábbi változata áll fenn. Tegyük fel, hogy $m > n$ és

$$\lim_{\substack{n\to\infty \\ m\to\infty}} n/m = \tilde{\lambda},$$

ahol $\tilde{\lambda}$ lehet 0 is. Ekkor $m, n \to \infty$ esetén

$$\tilde{T} = \sqrt{n}(U - \theta) \implies \mathcal{N}(0, \tilde{\lambda}\sigma_{10}^2 + \sigma_{01}^2).$$

(Amennyiben a határeloszlás szórása 0, akkor elfajult, pontosabban a 0-ba koncentrált, a határeloszlás.)

A bizonyítás ugyanaz, mint az 1.2. tételé, csak a $\sqrt{m}$ helyett $\sqrt{n}$-nel kell normálni.

6. Érvényben marad-e az 1.2. tétel, ha csupán $n \to \infty$, de m korlátos?

6. MEGOLDÁS. Igen. Ekkor $\lambda = 0$ és $\mathcal{N}(0, \sigma_{10}^2)$ a határeloszlás. Továbbá, ha $m \to \infty$, de n korlátos, akkor (lásd az 5. Feladatot) $\tilde{T} = \sqrt{n}(U - \theta) \implies \mathcal{N}(0, \sigma_{01}^2)$, hiszen $\tilde{\lambda} = 0$.

2. A Mann-Whitney-féle U-próba

2.1. A Mann-Whitney-féle U-próba bevezetése

A **Mann-Whitney-féle U-próbát** (más néven a **Wilcoxon-féle rangösszeg próbát**) mint speciális Hoeffding-féle statisztikát tekintjük. Legyen $X_1, \dots, X_m$ minta F eloszlásfüggvényű, míg $Y_1, \dots, Y_n$ minta G eloszlásfüggvényű sokaságból. Felteszük a két minta egymástól való függetlenségét is. Legyen az F és a G eloszlásfüggvény folytonos. F és G számunkra ismertelen, és éppen az F és G azonosságát kívánjuk vizsgálni.

Legyen

$$D_{ij} = \begin{cases} 1, & \text{ha } X_i > Y_j, \\ 0, & \text{ha } X_i \le Y_j, \end{cases}$$

$i = 1, \dots, m$, $j = 1, \dots, n$. Ekkor

$$U'_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} = \sum_{i=1}^{m} (\operatorname{rang}(X_i) - i)$$

a Wilcoxon-féle rangösszeg statisztika. Itt $\operatorname{rang}(X_i)$ az X_i sorszáma az $X_1, \dots, X_m$ és az $Y_1, \dots, Y_n$ minták együttes rendezése esetén.

2.1. tétel. *(A Wilcoxon-féle rangösszeg statisztika várható értéke és szórása.)*

$$\mathbb{E}U'_{mn} = mn\pi\,,$$

$$\mathbb{D}^2 U'_{mn} = nm\left[\pi - \pi^2(n+m-1) + (n-1)\pi_1 + (m-1)\pi_2\right]\,,$$

ahol

$$\pi = \mathbb{P}(Y < X) = \int_{-\infty}^{+\infty} G(x)dF(x) = \mathbb{E}G(X) = \mathbb{E}\big(1 - F(Y)\big),$$

$$\pi_1 = \mathbb{P}(Y_j < X_i, Y_k < X_i) = \int_{-\infty}^{+\infty} G^2(x)dF(x) = \mathbb{E}G^2(X),$$

$$\pi_2 = \mathbb{P}(Y_i < X_j, Y_i < X_k) = \int_{-\infty}^{+\infty} \big(1 - F(y)\big)^2 dG(y) = \mathbb{E}\big(1 - F(Y)\big)^2\,,$$

$j \neq k$. *Speciálisan,*

$$H_0 \ : \ F(x) \equiv G(x)$$

esetén

$$\mathbb{E}U'_{mn} = \frac{mn}{2}\,,$$

$$\mathbb{D}^2 U'_{mn} = \frac{nm(n+m+1)}{12}\,.$$

Bizonyítás. Határozzuk meg először U'_{mn} várható értékét! Mivel X és Y független,

$$\pi = \mathbb{P}(Y < X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{x} dG(y)dF(x) = \int_{-\infty}^{+\infty} G(x)dF(x) = \mathbb{E}G(X).$$

Ha $H_0 : F(x) \equiv G(x)$ teljesül, akkor

$$\pi = \int_{-\infty}^{+\infty} F(x)dF(x) = \mathbb{E}F(X) = \frac{1}{2},$$

lévén $F(X)$ a $[0,1]$-en egyenletes eloszlású (lásd az 1. Feladatot).

Másrészt nyilván $\mathbb{E}D_{ij} = \pi$, így $\mathbb{E}U'_{mn} = mn\pi$. Tehát ha H_0 teljesül, akkor $\mathbb{E}U'_{mn} = \frac{mn}{2}$.

Most $\mathbb{D}^2U'_{mn}$ összetevőit határozzuk meg. $\mathbb{D}^2D_{ij} = \pi(1-\pi)$ adódik D_{ij} definíciójából (hiszen Bernoulli-eloszlásról van szó). A függetlenség miatt

$$\operatorname{cov}(D_{ij}, D_{k\ell}) = 0, \qquad \text{ha} \quad i \neq k \quad \text{és } j \neq \ell.$$

A $\operatorname{cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ képlet alapján

$$\operatorname{cov}(D_{ij}, D_{ik}) = \pi_1 - \pi^2, \qquad \text{ha} \quad j \neq k,$$

ahol

$$\pi_1 = \mathbb{P}(Y_j < X_i, Y_k < X_i) = \int_{-\infty}^{+\infty} \mathbb{P}(Y_j < x, Y_k < x)dF(x) =$$

$$= \int_{-\infty}^{+\infty} G^2(x)dF(x) = \mathbb{E}G^2(X).$$

Hasonlóan

$$\operatorname{cov}(D_{ij}, D_{kj}) = \pi_2 - \pi^2, \qquad \text{ha } i \neq k,$$

ahol

$$\pi_2 = \mathbb{P}(X_i > Y_j, X_k > Y_j) = \int_{-\infty}^{+\infty} (1 - F(x))^2 dG(x).$$

Így

$$\mathbb{D}^2U'_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{D}^2(D_{ij}) + \sum_{j\neq k}\sum \operatorname{cov}(D_{ij}, D_{ik}) +$$

$$+ \sum_{i\neq k}\sum \operatorname{cov}(D_{ij}, D_{kj}) + \sum_{\substack{i\neq k \\ j\neq \ell}}\sum \operatorname{cov}(D_{ij}, D_{k\ell}) =$$

$$= nm\pi(1-\pi) + mn(n-1)(\pi_1 - \pi^2) + nm(m-1)(\pi_2 - \pi^2) =$$

$$= nm\left[\pi - \pi^2(n+m-1) + (n-1)\pi_1 + (m-1)\pi_2\right].$$

Speciálisan, $H_0 : F(x) \equiv G(x)$ esetén

$$\pi_1 = \int_{-\infty}^{+\infty} F^2(x)dF(x) = \mathbb{E}F(X)^2 = \frac{1}{3}, \qquad \pi_2 = \frac{1}{3},$$

ahol megint az $F(X)$ valószínűségi változó $[0,1]$-en való egyenletes eloszlását használtuk ki. Tehát H_0 esetén

$$\mathbb{D}^2U'_{mn} = \frac{nm(n+m+1)}{12}. \qquad \square$$

2.2. A Mann-Whitney-statisztika határeloszlása a Hoeffding-statisztika alapján

Láttuk, hogy

$$\varphi(x,y)=\begin{cases}1, & \text{ha } x>y,\\ 0, & \text{ha } x\le y\end{cases}$$

esetén $U'_{mn}=mnU$ éppen a Mann-Whitney-statisztika, ahol U a Hoeffding-statisztika. Ekkor (lásd a 3. Feladatot)

$$\sigma_{10}^2=\pi_1-\pi^2 \quad \text{és} \quad \sigma_{01}^2=\pi_2-\pi^2\,. \tag{2.1}$$

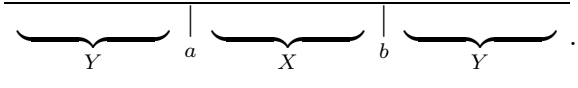
Tehát az 2.1 tétel alapján $\sigma_{10}^2=\mathbb{D}^2G(X)$ és $\sigma_{01}^2=\mathbb{D}^2\big(1-F(Y)\big)=\mathbb{D}^2F(Y)$.

Most már az 1.2 tétel segítségével meg tudjuk vizsgálni, hogy U'_{mn} mikor lesz aszimptotikusan normális, illetve mikor elfajult. Ha $\lambda>0$, úgy U'_{mn} akkor lesz elfajult (aszimptotikusan és véges m, n-re is), ha $\sigma_{10}^2=\sigma_{01}^2=0$.

$$0=\sigma_{10}^2=\mathbb{D}^2G(X) \iff \mathbb{P}(G(X)=\text{const.})=1 \iff$$

$$G(x)=c, \quad \text{ha} \quad x\in(a,b] \quad \text{és} \quad \mathbb{P}\big(X\in(a,b]\big)=1,$$

azaz X és Y eloszlása a számegyenesen az alábbi séma szerint alakul (ahol $a=\infty$, ill. $b=\infty$ is lehetséges)



$\sigma_{01}^2=0$ jelentése hasonló. Tehát

$$\sigma_{10}^2=0 \quad \text{és} \quad \sigma_{01}^2=0 \quad \Longleftrightarrow \quad \mathbb{P}(X\le Y)=1 \quad \text{vagy} \quad \mathbb{P}(X>Y)=1.$$

Az első esetben $U'_{mn}=0$, a másodikban $U'_{mn}=mn$, azaz értékük determinisztikus. Tehát, ha $\lambda>0$, ezen elfajult esetektől eltekintve U'_{mn} aszimptotikusan normális.

Ha H_0 teljesül, akkor ezek az elfajult esetek nem fordulnak elő.

2.2. tétel. *(A Wilcoxon-féle rangösszeg statisztika aszimptotikus normalitása.) Tegyük fel, hogy F és G folytonos, valamint H_0 : $F(x)\equiv G(x)$ teljesül. Ha $m,n\to\infty$ úgy, hogy $m/n\to\lambda$, akkor*

$$\frac{U'_{mn}-\frac{mn}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}$$

standardizált Wilcoxon-féle rangösszeg statisztika eloszlása a standard normális eloszláshoz konvergál.

Bizonyítás. Ekkor sem $\sigma_{10}^2=0$, sem $\sigma_{01}^2=0$ nem fordulhat elő. Tehát ekkor a 1.2 tételbeni a határeloszlás nem lehet elfajult. (Lásd az 5. Feladatot is!) □

A Wilcoxon-féle rangösszeg próba leírása megtalálható az alábbi helyeken: Gibbons (1971), 142-146. o., Lehmann–D'Abrea (1975), 20. o. Az aszimptotikus normalitás

bizonyítása Lehmann–D'Abrea (1975), 362. o., míg az alkalmazások Hollander–Wolfe (1973), 68. o. és Vincze–Varbanova (1993), 101. o. lelhetők.

2.3. Feladatok

1. Legyen a ξ eloszlásfüggvénye az $F(x)$ folytonos függvény. Lássuk be, hogy $F(\xi)$ egyenletes eloszlású a $[0,1]$-en!

1. MEGOLDÁS. Belátjuk, hogy $F(\xi)$ eloszlásfüggvénye éppen a $[0,1]$-en egyenletes eloszlásfüggvény. $\mathbb{P}(F(\xi) < y) = 0$, ha $y \leq 0$, és $\mathbb{P}(F(\xi) < y) = 1$, ha $y > 1$, nyilvánvaló. Legyen $0 < y \leq 1$ rögzített. Belátjuk, hogy

$$F(a) < y \Longleftrightarrow a < x_y,$$

ahol $x_y = \sup\{x : F(x) < y\}$. Ugyanis, ha $a < x_y$, akkor $a < x_0$ legalább egy x_0-ra az $\{x : F(x) < y\}$ halmazból. De akkor $F(a) \leq F(x_0) < y$. A másik irányhoz belátjuk, hogy $F(x_y) = y$. F monoton növekvő és folytonos voltából: $F(x_y) = \sup\{F(x) : F(x) < y\}$. De, szintén F folytonos voltából, ez utóbbi éppen y-nal egyenlő. Visszatérve a másik irányhoz: ha $F(a) < y$, akkor — az F monoton növekvő voltát is használva — $a < x_y$. Végül $F(\xi)$ eloszlásfüggvényére:

$$\mathbb{P}(\omega : F(\xi(\omega)) < y) = \mathbb{P}(\omega : \xi(\omega) < x_y) = F(x_y) = y,$$

ha $0 < y \leq 1$.

2. Legyen ξ egyenletes eloszlású a $[0,1]$-en. Lássuk be, hogy $1 - \xi$ is egyenletes eloszlású a $[0,1]$-en! Hol volt erre szükség?

2. ÚTMUTATÁS. Alkalmazzuk az eloszlásfüggvényt!

3. Lássuk be az (2.1) összefüggést!

3. MEGOLDÁS.

$$\begin{aligned}
\sigma_{10}^2 = \mathbb{D}^2\{\psi_{10}(X)\} &= \mathbb{D}^2\left\{(\mathbb{E}\psi(x,Y))_{|x=X}\right\} = \\
&= \mathbb{D}^2\left\{(\mathbb{E}\varphi(x,Y))_{|x=X}\right\} = \mathbb{D}^2\left\{(\mathbb{P}(Y<x))_{|x=X}\right\} = \\
&= \mathbb{D}^2\left\{(G(x))_{|x=X}\right\} = \mathbb{D}^2 G(X) = \mathbb{E}G^2(X) - (\mathbb{E}G(X))^2 = \\
&= \pi_1 - \pi^2.
\end{aligned}$$

4. A $H_0 : F(x) \equiv G(x)$ teljesülése esetén határozzuk meg U'_{mn} várható értékét és szórását elemi eszközökkel!

4. MEGOLDÁS. Ekkor X és Y független, azonos eloszlású, folytonos eloszlásfüggvényű. Tehát $\mathbb{P}(X = Y) = 0$, $\mathbb{P}(X < Y) = \mathbb{P}(X > Y)$. Azaz $\pi = \mathbb{P}(Y < X) = 1/2$. X_i, Y_j, Y_k is független, azonos eloszlású. Bármely sorrendjük egyformán valószínű. Tehát $\pi_1 = \mathbb{P}(Y_j < X_i, Y_k < X_i) = 1/3$.

5. Érvényben marad-e az 2.2. tétel, ha csupán n és m egyike tart végtelenhez, a másik pedig korlátos?

5. MEGOLDÁS. Igen. H_0 esetén sem $\sigma_{10}^2 = 0$, sem $\sigma_{01}^2 = 0$ nem fordul elő, így a határeloszlások nem elfajultak (lásd az előző szakasz 6. Feladatát).

Tárgymutató