Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Hálózati Rendszerek és Szolgáltatások Tanszék

Matematikai statisztika

Összefoglaló

Készítette
Papp Dorottya

2017. március 11.

Tartalomjegyzék

1. fejezet

Alapfogalmak

A *sokaság*, vagy populáció, a vizsgálat tárgyát képező nagyszámú, de véges elemszámú egyedek halmaza. A statisztika, mint tudomány célja a halmaz egészének kevés adattal történő tömör jellemzése és az egyedek leírására a bevezetett változók közötti kapcsolatok leírása. Nincs lehetőség (erőforrás), hogy a populáció minden egyes eleméről adatokat szerezzünk be. Példa sokaságok: ország állampolgárai, gyártott termékek. A statisztikai elemzés tárgya lehet egy *véletlen kísérlet* is, ami időben változatlan körülmények között elvileg akárhányszor lejátszódhat, pl. terméshozam, lottóhúzás, véletlen egyed kiválasztása.

A statisztikai minta *realizáltja* a populáció egy kis elemszámú részhalmazára vonatkozó megfigyelések adatai. Például: hőmérséklet adatok, felmérésbe bevont nézők. A mintának reprezentatívnak kell lennie, vagyis tükröznie kell a populáció tulajdonságait. A populáció minden egyedének ugyanakkora esélyt kell biztosítani a mintába kerüléshez, a minta elemszámának pedig elég nagynak kell lennie ahhoz, hogy a következtetések átvihetők legyenek a populációra is. A mintavételezésnek 3 nagy kategóriája:

- Rétegzett mintavételezés: populációt adott szempontok szerint csoportokba osztjuk és a csoportok arányait a mintában is megtartjuk
- Véletlen mintavételezés: mintába kerülő egyedeket sorsolással választjuk ki
- Cenzus: népszámlálás

A mintát egy adatmátrixba írhatjuk fel, aminek n sora és p oszlopa van. Minden sor megfelel a minta egy elemének, minden oszlop pedig egy *változónak*, a populáció egy mérhető jellemzőjének.

A *statisztika* a minta realizáció adataiból adott képlettel számolt adat. Pl. átlag, medián, gyakoriság, stb.

1.1. Alapmodell

A matematikai statisztika alapmodellje 4 elemből áll:

- $\mathscr{K}$, a véletlen kísérlet,
- Ω, a lehetséges kimenetelek halmaza,
- $\mathscr{A}$, a megfigyelhető események halmaza, és

- $\mathscr{P}$, a lehetséges valószínűségi mértékek halmaza.

Az elemzés célja, hogy $\mathscr{P}$-ből kiválasszuk a tényleges valószínűséget, de legalább is egy jó helyettesítő egyedet.

A matematikai alapmodellben a változó $X : \Omega \to \mathbb{R}$ egy valószínűségi változó, melynek minden $\mathbf{P} \in \mathscr{P}$ esetén megadható az eloszlásfüggvénye: $F_x(t) = \mathbf{P}(X < t)$. Ekkor a feladat a következő halmazból kiválasztani a vlóságot legjobban leíró eloszlásfüggvényt:

$\mathscr{F} = \{F_x(t) : F_x(t) = \mathbf{P}(X < t) : \forall \mathbf{P} \in \mathscr{F}\}$

A statisztikai minta ez alapján az X valószínűségi változóval azonos eloszlású, egymással teljesen független $X_1, X_2, ...X_n$ valószínűségi változók együttese. Egy mintavételkor tulajdonképpem megfigyeljük a $\mathscr{K}$ véletlen kísérletet, aza megállapíTjük, hogy melyik $\omega \in \Omega$ kimenetel realizálódott. Az $X_1(\omega) = x_1, ..., X_n(\omega) = x_n$ szám n-est nevezzük a minta realizációjának.

A statisztika egy valószínűségi változó, aminek eloszlásfüggvényét a minta eloszlásfüggvényéből lehet kiszámolni: $T_n = t_n(X_1, X_2, ..., X_n)$. A statisztika számolt értéke az, amikor $t_n()$ argumentumába a mintarealizáció értékeit helyettesítjük.

Adatcentrum statisztikái:

- átlag: $\displaystyle\frac{1}{n}\sum_{i=1}^{n} x_i$
- medián: $m_n = \begin{cases} x_{\frac{n+1}{2}} & \text{ha } n \text{ páratlan} \\ \frac{x^*_{\frac{n}{2}} + x^*_{\frac{n+1}{2}}}{2} & \text{ha } n \text{ páros} \end{cases}$
- módusz: leggyakrabban előforduló érték a mintában
-

Szóródást jellemző statisztikák:

- standard szórás: $s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, ahol $\bar{x}$ az átlag
- szórás torzítatlan becsléssel: $s^*_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- variáció: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- terjedelem: $x^*_n - x^*_1$ (legkisebb és legnagyobb elem közötti különbség)

Eloszlást jellemző statisztikák:

- ferdeség: $s = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^3}$, ha $s > 0$, akkor jobbra ferdül, ha $s < 0$, akkor balra ferdül
- lapultság: $k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^4} - 3$, ha $k > 0$, akkor csúcsos, ha $k < 0$, akkor lapos.

A mintaelemeket szokás rendezni is: $X_k^* = ord_k(X_1, X_2, ..., X_n), k = 1, 2, ..., n$[1]. Ezeket rendezett mintaelem-statisztikáknak hívjuk. A minta eloszlásfüggvénye és a rendezett minta eloszlásfüggvénye között az alábbi összefüggés áll fenn:

$F_k(x) = \mathbf{P}(X_k^* < x) = \sum_{i=k}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i}$

A rendezett mintaelem-statisztikák közül különlegesen fontos az empirikus eloszlásfüggvény:

$$F_n(x) = \begin{cases} 0 & \text{ha } x \leq X_1^* \\ \frac{k}{n} & \text{ha } X_k^* < x \leq X_{k+1}^* \\ 1 & \text{ha } x > X_n^* \end{cases}$$

Fontosságát a matematikai statisztika alaptétele adja:

$\mathbf{P}\Big(\lim_{n \to \inf} sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\Big) = 1$, vagyis az empirikus eloszlásfüggvény 1 valószínűséggel, egyenletesen konvergál az elolszlásfüggvényhez.

1.2. Fontos eloszlások

- χ^2 eloszlás
- Student-eloszlás
- F-eloszlás

[1] $ord()$ kiválasztja a k-dik legnagyobb elemet az argumentumok közül

2. fejezet

Statisztikai becslések

Előfordul, hogy a minta eloszlásfüggvénye egy képlet, amelyet a ϑ paraméter konkretizál. Ha ismerjük az értékét, meg tudjuk pontosan adni az eloszlásfüggvényt:
$\mathscr{F} = \{F_x(t, \vartheta) : \vartheta \in \Theta\}$

Egy adott statisztikai minta segítségével a ϑ paraméter megbecslése a célunk. A paraméter becsléséhez valamilyen alkalmas T_n statisztikát használunk, egy ismeretlen számot egy valószínűségi változóval becsülünk. Ekkor persze felvetődik a kérdés: mikor jó egy ilyen becslés?

- Torzítatlanság: $\mathbf{E}T_n = \vartheta$, vagyis a becslő statisztika pont a becsülendő paraméterérték körül fogja felvenni az értékeit. Indok: egy valószínűségi változó az összes szám körül épp a várható érték körül ingadozik a legkisebb mértékben
  - Aszimptotikus torzítatlanság: a torzítatlansági feltétel csak $n \to \infty$ esetében igaz: $\lim_{n\to\infty} \mathbf{E}T_n = \vartheta$
  - A Cramer-Rao egyenlőtlenség elvi alsó korlátot ad a torzítatlan becslések szórásnégyzeteire. Ha egy statisztikára belátjuk, hogy a szórásnégyzete éppen az alsó korláttal egyenlő, akkor az biztosan hatásos (amiből egyetlen egy van)
- Konzisztencia: $\forall\epsilon > 0$ szám esetén teljesül, hogy $\lim_{n\to\infty} \mathbf{P}(|T_n - \vartheta| < \epsilon) = 0$, vagyis garancia van arra, hogy a minta elemszám növekedtével növekszik a becslés pontosságának valószínűsége
  - Erős konzisztencia: $\mathbf{E}T_n = \vartheta$ és $\lim_{n\to\infty} \mathbf{D}^2 T_n = 0$, vagyis azok a torzítatlan becslések, ahol az elemszám növekedtével a szórásnégyzet (variancia) 0-hoz tart. Az erősen konzisztenc statisztikai becslések egyben konzisztensek is
- Hatásosság: két torzítatlan becslés közül a kisebb varianciájú a jobb, hiszen kisebb mértékben ingadozik a paraméter körül, vagyis kevesebb megfigyeléssel is jó becslés kapható. Egy torzítatlan becslés akkor lesz *hatásos*, ha varianciája minden más torzítatlan becslés varianciájánál kisebb. Hatásos becslésből egyetlen egy létezik csak, ezt érdemes megkeresni egy adott paraméter-becslési problémához
- Elégségesség: egymaga képes helyettesíteni a mintát, a paraméterre vonatkozóan minden információt magába sűrít

  - Rao-Blackwell-Kolmogorov tétel: ha létezik hatásos (legjobb torzítatlan) becslés, akkor elég azt az elégséges becslés függvényei között keresni
  - Ellenőrzini a Neymann-Fisher faktorizációs tétellel lehet

Jó tudni:

- Az átlagstatisztika a várható értéknek, mint paraméternek, a torzítatlan becslése (ha létezik a minta szórása, akkor erősen konzisztens), normális, exponenciális és Poisson esetben hatásos
- A tapasztalati szórásnégyzet a minta varianciájának aszimptotikusan torzítatlan becslése (ha létezik a minta negyedik momentuma, akkor konzisztens is)
- A korrigált empirikus szórásnégyzet statisztik a variancia torzítatlan becslése (ha létezik a minta negyedik momentuma, akkor erősen konzisztens), normális esetben hatásos

2.1. Maximum likelihood becslés

A módszer alapgondolata, hogy ha egy kísérletnél több esemény is bekövetkezhet, akkor legtöbbször a legnagyobb valószínűségű esetményt fogjuk megfigyelni. A mintavételezés során kaptunk egy realizációt, errőr a realizációról pedig az előzőek miatt feltételezzük, hogy azért ezt kaptuk, mert ennek volt a legnagyobb bekövetkezési valószínűségee. Vagyis, az összes lehetséges ϑ paraméter (amitől függ a minta eloszlásfüggvénye) közül vegyük azt, amelynél a kapott realizáció bekövetkezése a maximális. Általános feltételed mellett a becslés konzisztens, aszimptotikusan normális eloszlású és ha van elégséges statisztikai, akkor éppen azt adja meg.

Számítása diszkrét esetben:

$L(\mathbf{x},\vartheta) = P_\vartheta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \sum_{i=1}^n P_\vartheta(X_i = x_i)$, a minta együttes eloszlása. A paraméter maximum likelhood becslése az a $\tau_n(X_1, X_2, ..., X_n)$ statisztika, melyre $L(x, \tau_n(\mathbf{x})) = \max_\vartheta L(\mathbf{x}, \vartheta)$

Számítása folytonos esetben:

$L(\mathbf{x},\vartheta) = P_\vartheta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \prod_{i=1}^n f_\vartheta(x_i)$, a minta együttes eloszlása. A paraméter maximum likelhood becslése az a $\tau_n(X_1, X_2, ..., X_n)$ statisztika, melyre $L(x, \tau_n(\mathbf{x})) = \max_\vartheta L(\mathbf{x}, \vartheta)$

A maximum meghatározásához $L(\mathbf{x},\vartheta)$ ϑ szerinti első deriváltjának zérushelyeit keressük. A zérushelyek közül az a maximum, ahol a második derivált negatív. Megjegyzés: ha $L(\mathbf{x},\vartheta)$ deriválása túl nehéz, érdemes a logaritmusát venni (loglikehood függvény, $l(\mathbf{x},\vartheta)$) és azzal számolni. Mivel a logaritmus függvény szigorúan monoton nő, ezért nem torzítja a zérushelyeket.

Többparaméteres esetben paraméterenként vizsgáljuk az első deriváltakat és a Hesse-mátrix alapján döntünk a maximum helyről. Maximum esetén a mátrix diagonális és az első diagonál elem negatív.

2.2. A momentumok módszere

Legyen adott a valószínűségi mértékek egy tere és az $X_1, X_2, ..., X_n$ statisztikai minta. Tegyük fel, hogy létezik az első k momentum: $m_j = \mathbf{E}_\vartheta X_i^j = g_j(\vartheta)$, vagyis a momentumok a paraméter valamilyen $g()$ függvényeként állnak elő. Tegyük fel, hogy $\exists g_j^{-1}(m_1, m_2, ..., m_k) = \vartheta_j$. Tekintsük az $\hat{m}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$ empirikus momentum statisztikákat. Ekkor a ϑ_j paraméteres momentumos becslései a $m_j = g_j^{-1}(\hat{m}_1, \hat{m}_2, ..., \hat{m}_k)$

2.3. Intervallumbecslések

Eddig az ismeretlen paramétervektort a minta egy függvényével, azaz egyetlen statisztikával próbáltuk meg közelíteni. Konkrét realizációnál tehát, a paramétertér egy pontját egy másik ponttal becsüljük (*pontbecslés*).

Folytonos eloszlásoknál azonban annak valószínősége, hogy a valószínőségi változó az értékkészletének éppen egy tetszőlegesen kiválasztott pontját fogja felvenni, nulla. Tehát folytonos esetben nulla annak valószínősége, hogy éppen a paramétert találtuk el a becsléssel. Az intervallumbecsléseknél a mintából készített tartományokat definiálunk, amely tartományok nagy valószínőséggel lefedik a kérdéses paraméterpontot.

Legyen adott a valószínésgi mértékeke egy tere és az $X_1, X_2, ..., X_n$ statisztikai minta. Legyen $0 < \epsilon < 1$ rögzített. A ϑ paraméterhez megadhatunk egy legalább $1 - \epsilon$ szignifikanciaszintű konfidencia-intervallumot, ha $t_1(X_1, X_2, ..., X_n)$ és $t_2(X_1, X_2, ..., X_n)$ olyan statisztikák, hogy:

$\mathbf{P}_\vartheta(t_1(X_1, X_2, ..., X_n) \leq \vartheta \leq t_2(X_1, X_2, ..., X_n)) \geq 1 - \epsilon$ mindig igaz.

3. fejezet

Hipotéziselmélet

Adott a $\mathscr{K}$ véletlen kísérlet, az Ω eseménytér, a lehetséges valószínűségek halmaza $\mathscr{P}$ és a felette értelmezett $\mathbf{P}$ Kolmogorov-féle valószínűségi mező. Tegyük fel, hogy $\mathscr{P}$ két diszjunkt halmazra bontható: $\mathscr{P}_0$ és $\mathscr{P}_1$. Statisztikai próbát akarunk kidolgozni annak eldöntésére, hogy $H_0 : \mathbf{P} \in \mathscr{P}_0$ *nullhipotézis* igaz-e. Ha úgy kell döntenünk, hogy anull hipotézis nem igaz, akkor automatikusan a $H_1 : \mathbf{P} \in \mathscr{P}_1$ *alternatív hipotézist* fogjuk elfogadni. A döntéshez szignifikancia szintet rendelünk, amivel jellemezzük, hogy mennyire erős a döntésünk.

Azonban a döntésünk nem lesz minden esetben helyes, elképzelhető, hogy hibásan döntünk. *Elsőfajú* hibáról beszélünk, ha H_1-et fogadjuk el, minközben H_0 igaz. Ennek értékét tudjuk befolyásolni, általában 5-10%-ra választjuk. A hibánk *másodfajú*, ha H_0-t fogadjuk el, miközben H_1 az igaz. A másodfajú hiba értékét nehéz megállítani. A kéz hibafajta között trade-offot kell találni, ha az egyiket csökkentjük, a másik nőni fog. A hibavalószínáségeket csak úgy tudjuk csökkenteni, ha növeljük a mintaelemszámot (mivel így csökken a szórás)!

3.1. Paraméteres próbák

Paraméteres próbák esetén feltételezzük, hogy $\mathscr{P} = \{\mathbf{P}_\vartheta : \vartheta \in \Theta\}$, ahol Θ a paramétertér, melyet két diszjunkt halmazra osztunk(Θ_0 és Θ_1). A nullhiptézisünk ez esetben, hogy $\vartheta \in \Theta_0$, az alternatív hipotézis pedig, hogy $\vartheta \in \Theta_1$.

A pontos döntéshez a minta eloszlásfüggvényét $F_\vartheta(x)$ közelítjük a $T_n(X_1, X_2, ..., X_n)$ próbastatisztikával. A szignifikacia szintet ε-nak jelöljük és feltesszük, hogy $\forall \varepsilon \in (0,1) : \exists K_1(\varepsilon) < \exists K_2(\varepsilon) : \mathbf{P}_\vartheta(K_1(\varepsilon) < T_n < K_2(\varepsilon)) \geq 1 - \varepsilon$. Ekkor az elfogadási tartomány: $\mathscr{X}_e = \underline{x} \in \mathscr{R}^n : K_1(\varepsilon) < T_n(\underline{x}) < K_2(\varepsilon)$, ahol $\underline{x}$ a mintarealizáció. Ha $\underline{x} \in \mathscr{X}_e$, akkor a nullhipotézist elfogadjuk az ε szignifikancia szinten.

A tanult paraméteres próbákban közös, hogy **az elemzett minta normális eloszlást követ**. A nullhiptézist a normális eloszlás paramétereivel (várható érték, szórás) fogalmazzuk meg. A próbák adatait a 3.1. táblázat tartalmazza. Mivel képletgyűjtemény használható, a képleteket kihagytam a táblázatból, csak az alkalmazáshoz szükséges információkat gyűjtöttem itt össze.

3.2. Nemparaméteres próbák

Nemparaméteres próbák esetén a statisztikai minta eloszlását nem tekintjük eleve ismertnek. Az előzetes feltevésink nagyon általánosak, pl. folytonos eloszlást követ

3.1. **táblázat.** *Paraméteres próbák*

Próba neve	**Feltétel**	Döntés
egymintás u-próba	szórás ismert	$\lvert u_{prba}\rvert < K_{krit}$
kétmintás u-próba	független statisztikai minták, szórásaik ismertek	$\lvert u_{prba}\rvert < K_{krit}$
egymintás t-próba	szórás nem ismert	$\lvert t_{prba}\rvert < K_{krit}$
független kétmintás t-próba	minták függetlenek, szórásai egyenlőeknek tekintendők ($\rightarrow$ F-próba)	$\lvert t_{prba}\rvert < K_{krit}$
összetartozó kétmintás t-próba		$\lvert t_{prba}\rvert < K_{krit}$
Welch-próba	független, normális eloszlású minták, eltérő szórással (amik nem ismertek)	$\lvert W_{prba}\rvert < K_{krit}$
F-próba	független, normális eloszlású minták, szórás nem ismert	tört $\in F_{n-1,m-1}$

minta, vagy véges a szórás. Mivel kevesebb feltételt követelünk meg kiinduláskor, nagyonn mintaelemszámra van szükségünk, mint a paraméteres próbák esetén.

Illeszkedés vizsgálat

Illeszkedés vizsgálat esetén azt vizsgáljuk, hogy az elemzett változó eloszlása megegyezik-e a hipotetikussal. Kapcsoló próbák: χ^2-próba, egymintás Kolmogorov-Szmirnov próba.

χ^2 próba (tiszta illeszkedésvizsgálat) esetén a minta értékkészletét fel kell bontanunk r diszjunk intervallumra. Ökölszabály: minden intervallumba kb. ugyanannyi elemnek kell esnie (θ_i)! Minden intervallumra meg tudjuk mondani a bekövetkezés gyakoriságát (p_i). A kritikus érték a χ^2 táblázatból jön.

Egymintás Kolmogoroc-Szmirnov mintánál az empirikus eloszlásfüggvénnyel $F_n(x)$ közelítjük a hipotetikus eloszlás függvényt $F_o(x)$. Kritikus értéket a Kolmogorov táblázatból nézhetünk.

Függetlenség vizsgálat

Azt vizsgáljuk, hogy az elemzett változók függetlennek tekinthetőek-e. Kapcsolódó próba: χ^2 próba nominális vagy ordinális változókra.

χ^2 próba (kétdimentiós statisztikai minta) esetén a két mintát külön-külön intervallumokra osztjuk, majd az intervallumokból táblázatot készítünk. Itt is érvényes az 1D-s ökölszabály kiterjeszése: minden kitöltött cellában kb. azonos nagysegrendbe szerepeljenek értékek! A táblázat celláinak értéke θ_{ij}, az adott sorhoz tartozó bekövetkezési valószínűség $p_i = \sum_{j=1}^{s} p_{ij}$, az oszlophoz $p_j = \sum_{i=1}^{r} p_{ij}$. A sorok száma r, az oszlopok száma s.

Homogenitás vizsgálat

Azt vizsgáljük, hogy az elemzett változók eloszlása azonos-e, vagyis az eloszlásfüggvények közösek-e. Kapcsolódó próbák: χ^2 próba, kétmintás Kolmogorov-Szmirnov próba, Mann-Whitney próba, Wilcoxon próba, Kruskal-Wallis próba, Friedmann próba.

χ^2 próba esetén két mintarealizációnk van: X elemszáma n_1, Y-é n_2. Az X mintarealizációt r intervallumra osztjuk. Összeszámoljuk, hogy hány minta esett egy-egy intervallumba az X mintarealizációból (θ_i) és mennyi az Y mintarealizációból λ_i. Kritikus értéket a χ^2 táblázatból tudunk nézni.

Kétmintás Kolmogorov-Szmirnov próba esetén feltételezzük, hogy a minták függetlenek különböző eloszlásfüggvénnyel. A homogenitás vizsgálatához a minták empirikus eloszlásfüggvényeinek ($F_{n_1}(x)$ és $(F)_{n_2}(x)$ különbségét vizsgáljuk. Kritikus értéket a Kolmogorov táblázatban találunk.

Mann-Whitney próba esetén az X minta adatait két részre osztjuk egy Y csoportképző változó segítségével, így kapjuk x (elemszáma n) és y (elemszáma m) részmintákat. A két minta összefésüléséből képezzük a $z_1^* \leq z_2^* \leq ... \leq z_{n+m}^*$ rendezett mintát. A rendezett minta elemeihez rangszámokat (r_i) rendelünk: a legkisebb elem rangszáma 1, a legnagyobbé $N = n + m$. Kritikus értéket a standard normális táblázatból kapjuk.

Kruskal-Wallis próba esetén több független minta együttes homogenitás-vizsgálatát szeretnénk elvégezni. A nullhipotézisünk, hogy a p független minta ugyanabból az eloszlásból származik-e. A p független mintát és Y tördelő változó segítségéven állítjuk elő az X változó eseteiből, vagyis a Y változó értékei p csoportra osztják az X minta eseteit (n_j a csoport elemszáma, N az összes csoport elemszámának összege). Ismét képezzük a rendezett mintát és az alapján az elemeket rangszámokkal látjuk el (R_j). Kritikus értéket a χ^2 táblázatban találunk.

Wilcoxon próba esetén azt ellenőrizzük, hogy az összetartozó mintákból készített adatmátrix X és Y változója azonos eloszlásfüggvényhez tartozik-e. Képezzük az összetartozó adatpárok különbségét (d_i), vegyük annak előjelét (s_i) és az eltérések abszolút értékét (a_i). CSináljunk az abszolút eltérésekből rendezett mintát (a_i^*) és ezekhez rendeljünk rangszámot. A pozitív differenciák rangszám-összege R_+. A kritikus értéket a standard normális táblázatban találjuk ($\Phi(u_\varepsilon) = 1 - \frac{\varepsilon}{2}$).

Friedmann próba esetén p változó azonos eloszláshoz tartozását ellenőrizzük. Az adatmátrix minden sorához rangszámokat rendelünk, ahol a rangszám megadja, hogy az adott elem hagyadik legkisebb az adott sorban. $R^{(j)}$ az egyes oszlopokhoz tartozó rangszám-összeg. Kritikus értéket a χ^2 táblázatban találunk, ha kicsi az elemszám.