VÉRTES LÁSZLÓ MATEMATIKAI STATISZTIKAI MUNKÁSSÁGÁRÓL

BOSZNAY ÁDÁM

Intersoft Informatika Kft.
Email: bosznay@intersoft.hu

Abstract

The paper describes some aspects of the application of mathematical statistical methods in the publications of L. Vértes. Demontrates the fact that Vértes used the methods of mathematical statistics to obtain archeological consequences. Shows also, that Vértes – though he was not a mathematician- used these methods with the rigorousness of a professional statistician.

Kivonat

A dolgozat a matematikai statisztikai módszerek alkalmazásának néhány példájával foglalkozik Vértes László munkássága köréből. Vértes alkotóan alkalmazta a matematikai statisztika eszköztárát régészeti eredmények, következtetések levonására. Bár Vértes László nem volt matematikus, az alkalmazott módszereket nagy körültekintéssel és hozzáértéssel használta fel munkáiban.

KEYWORDS: MATHEMATICAL STATISTICS, STUDENT-TEST, ARCHAEOMETRY

KULCSSZAVAK: MATEMATIKAI STATISZTIKA, T-PRÓBA, ARCHEOMETRIA

Bevezetés

Jelen dolgozat Vértes László matematikai statisztikai munkásságából kíván egy-két részletet bemutatni. Semmiképpen sem törekszik a teljességre. Az itt ismertetésre kerülő matematikai statisztikai eljárások részletesebb ismertetését illetően lásd pl. Vincze (1968) művét.

Alapfogalmak

Vértes László munkáiban számos, a matematikai statisztikában jól ismert és elfogadott módszert és fogalmat használt. A legfontosabb fogalmakat - Vértes gondolatmenetének jobb megismerése okán - az alábbiakban foglalhatjuk össze.

A dolgozatban említésre kerülnek a normális és lognormális valószínűségeloszlások. Ezek fontosabb tulajdonságait illetően Prékopa (1962) és Rényi (1966) műveire utalunk, röviden azonban bevezetjük e fogalmakat,s utalunk néhány tulajdonságukra – nem a teljes pontosság igényével.

Egy valószínűségeloszlást normálisnak nevezünk, ha tetszőleges x értékre annak valószínűsége, hogy az eloszlás értéke kisebb, mint x, az alábbi formulával határozható meg.

$$\phi(x,\mu, \sigma)=\int^{x}_{-\infty} 1/\sigma\ 1/(2\Pi)^{½} \exp(-(y-\mu)^2/2)\ dy$$

Itt a $\sigma > 0$ paraméter az eloszlás szórása, míg a μ paraméter annak várható értéke.

A fenti integrálandó kifejezést szokták a normális eloszlás sűrűségfüggvényének nevezni.

Egy valószínűségeloszlást lognormálisnak nevezünk, ha logaritmusa normális eloszlású.

Néhány példát láthatunk normális eloszlás-függvényekre az 1., sűrűségfüggvényekre a 2. ábrán.

Közelítőleg normális eloszlást követnek azok az értékek, melyek (minél) több egymástól független véletlen, összeadódó hatás átlagaként származtathatóak.

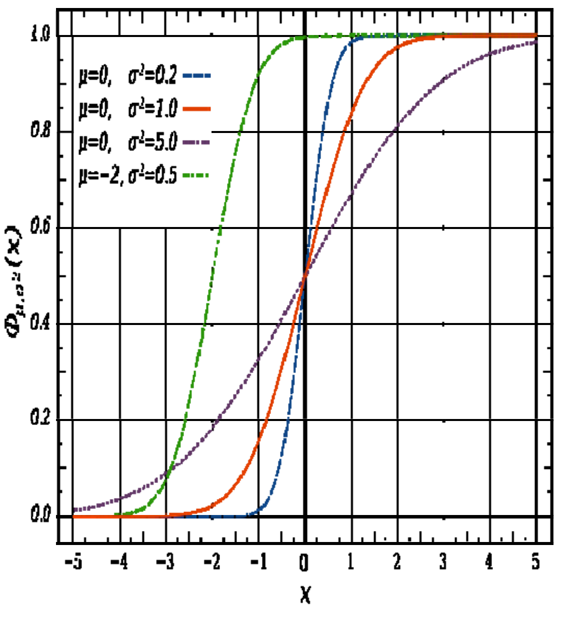
Néhány példát láthatunk lognormális eloszlásfüggvényekre a 3., sűrűségfüggvényekre a 4. ábrán.

Egy valószínűségeloszlást csonkolt normálisnak nevezünk a $[t_1,t_2]$ intervallumra nézve, ha tetszőleges $t_1 \leq x \leq t_2$ értékre nézve annak valószínűsége, hogy az eloszlás értéke kisebb, mint x, az alábbi formulával határozható meg.

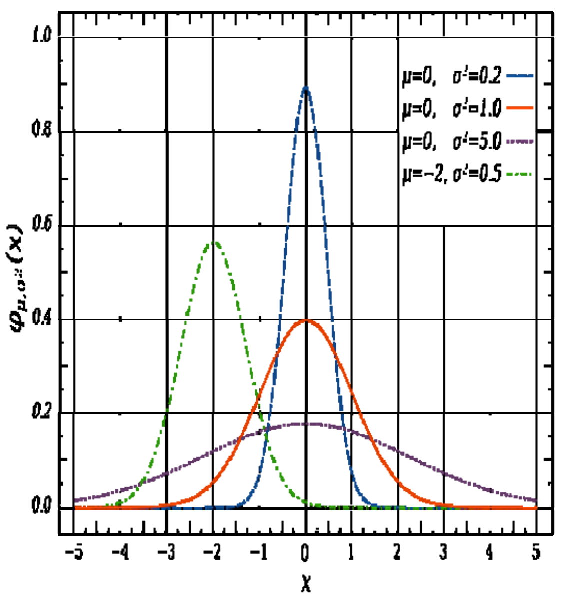
$$(\phi(x, \mu, \sigma) - \phi(t_1, \mu, \sigma))/(\phi(t_2, \mu, \sigma) - \phi(t_2, \mu, \sigma)).$$

Szemléletesen ez azt jelenti, hogy az eredeti eloszlásból kihagyjuk azokat az értékeket, amelyek a $[t_1, t_2]$ intervallumon kívül esnek. Ilyen eloszláshoz nyilván akkor jutunk, ha a „fel” vagy „le” kiugró értékeket egy normális eloszlásból elhagyunk.

Közelítőleg lognormális eloszlást követnek azok az értékek, melyek (minél) több, egymástól független véletlen, szorzódó hatásból származtathatóak. Így pl. egy véletlenszerűn végzett kőzetdaraboláskor a kapott darabok méreteloszlása jó közelítéssel lognormális.



1. ábra - normális eloszlásfüggvények



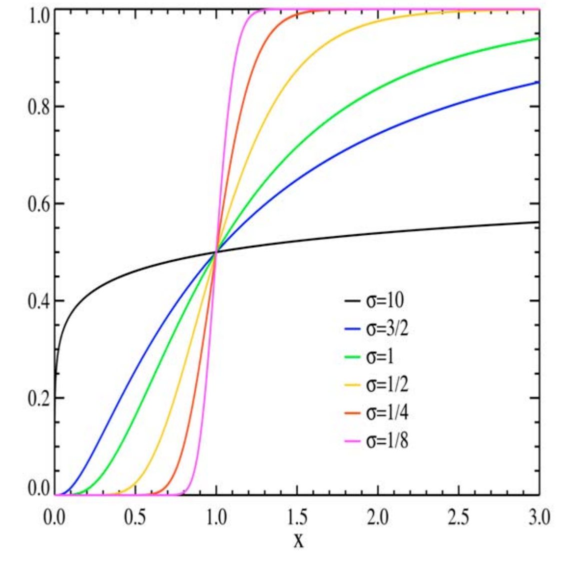
2. ábra - normális sűrűségfüggvények

K szabadságfokkal rendelkező Khi^2 eloszlást kapunk akkor, ha k darab 0 várható értékű és 1 szórású, független normális eloszlású valószínűségi változó négyzetösszegét képezzük.

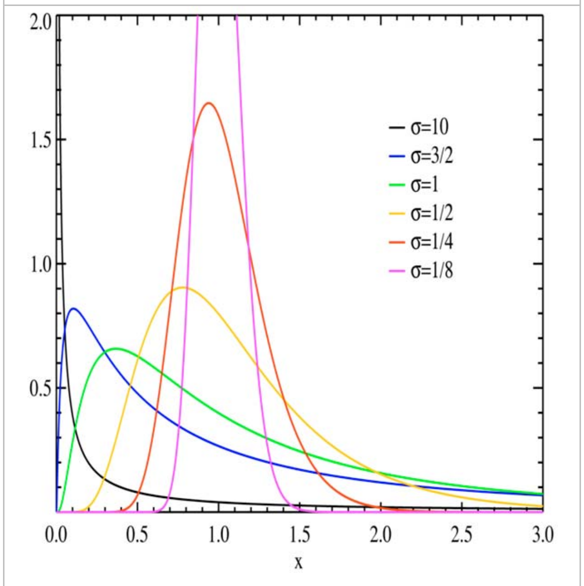
K szabadságfokkal rendelkező T-eloszlást kapunk, ha tekintünk egy 0 várható értékű és 1 szórású normális eloszlású változót, és egy, tőle független k szabadságfokú Khi^2- eloszlású H változót, s képezzük az alábbi kifejezést:

$$N/(H/(k-1))^{1/2}.$$

Az alábbiakban formulákkal mutatjuk be azokat a statisztikai próbákat, melyeket Vértes alkalmazott.



3. ábra - lognormális eloszlásfüggvények



4. ábra - lognormális sűrűségfüggvények

A próbák megnevezése mögött közöljük azok Vértes által használt funkcióját is. Mielőtt ezt megtennénk, néhány jelölést vezetünk be.

Legyen $x_1,\ldots x_{nx}$ egy, nx elemű minta. Ekkor jelölje $\bar{x}$ e mintaelemek átlagát, azaz

$$\bar{x} = (x_1+\ldots+x_{nx})/nx.$$

Egy minta terjedelmének szokás nevezni legnagyobb és legkisebb értékének különbségét.

Az alábbi értéket a minta korrigált tapasztalati szórásnégyzetének nevezzük.

$$s^2_x = ((x_1-\bar{x})^2+\ldots+(x_{nx}-\bar{x})^2)/(nx-1)$$

Megjegyezzük, hogy valamely eloszlás várható értékének becslésére a megfelelő eloszlásból származó minta átlagát, míg szórásának becslésére

HU ISSN 1786-271X; urn:nbn:hu-4106

a tapasztalati szórásnégyzet négyzetgyökét célszerű használni.

T-próba: Funkciója annak eldöntése, hogy két, azonos szórással bíró normális eloszlású minta várható értéke azonos-e.

A megfelelő formula

$$(\bar{x} - \bar{y})/s_{xy}/(1/nx+1/ny),$$

ahol

$$s_{xy}^2 = ((nx-1)/\, s^2_x + (ny-1)/\, s^2_y)/(nx+ny-2).$$

A megfelelő T-eloszlás szabadságfoka nx+ny-2, ez a kívánt (T) táblázati sor kiválasztásához szükséges.

Welch-próba: Funkciója annak eldöntése, hogy két, nem feltétlenül azonos szórással bíró normális eloszlású minta várható értéke azonos-e.

A megfelelő formula

$$(\bar{x} - \bar{y})/(\, s^2_x /nx + s^2_y /ny)^{\,1/2},$$

amely kifejezés a várható értékek azonossága és mindkét minta elég nagy elemszáma esetén közelítően normális eloszlású, 0 várható értékkel és 1 szórással.

F próba: Funkciója annak eldöntése, hogy két normális eloszlású minta szórása megegyezik-e.

A megfelelő formula

$$s^2_x / s^2_y$$

amely a hipotézis fennállása esetén (nx-1,ny-1) paraméterű ún. F-eloszlást követ.

Bartlett-próba: Funkciója annak eldöntése, hogy több normális eloszlású minta szórása azonos-e.

Legyenek a normális eloszlású minták darabszámai $n_1,...n_k$. Legyen továbbá $N= n_1+...+n_k$

Legyenek a megfelelő korrigált tapasztalati szórásnégyzetek $s^2_1,...s^2_k$, továbbá

$$s^2_p = 1/(N-k) *((n_1-1)\, s^2_1 + ... + (n_k-1)\, s^2_k).$$

Ekkor az alábbi kifejezés

$$((N-k)\ln(s^2_p) - (n_1-1)\ln(s^2_1) - ... - (n_k-1)\ln(s^2_k))/(1 + 1/(3(k-1))*(1/(n_1-1)+...+ 1/(n_k-1)-1/(N-k))$$

ahol ln() a természetes logaritmus függvényt jelöli; k-1 szabadságfokú Khi^2 eloszlású valószínűségi változó.

Khi^2 próba: Funkciója annak eldöntése, hogy két minta eloszlása azonos-e.

Ennek elvégzéséhez mindkét mintát osztályokba kell sorolni. Ez azt jelenti, hogy tekinteni kell k darab olyan közös elem nélküli osztályt, mely mindkét minta összes elemét tartalmazza. Az i. osztályba eső mintaelemek számát az x mintában jelölje nx_i, az y mintában pedig ny_i.

A megfelelő formula

$$nx*ny\ *((\, ny_1/ny - nx_1/nx)^2 \ /((nx_1+ny_1)+...+ ((\, ny_k/ny - nx_k/nx)^2 \ /((nx_k+ny_k))$$

a hipotézis fennállása esetén közelítőleg Khi^2 – eloszlású, k-1 szabadságfokkal.

Matematikai statisztikai alkalmazások Vértes László munkáiban

Vértes (1964) statisztikai módszerei és következtetései az alábbiak.

Először 2 ásatásból (1960/61 és 1964) származó darabok méreteloszlását veti össze az ugyanezen a helyen, felszínen, gyűjtők által összeszedett kb. 300 kőeszközzel. A két ásatás eredményét közös mintába, míg a „gyűjtöttet” más mintába csoportosítva, e két minta eloszlását hasonlítja össze Khi^2 – próbával. Eredményként adódik, hogy a két eloszlás nem azonos. Ezt annak a ténynek tudja be, hogy az ilyen gyűjtés esetén a gyűjtő feltehetően forma-prekoncepcióval válogat. Ez óvatosságra intheti a szakembert a nem szakember által gyűjtött minta felhasználhatóságát illetően.

Arka lakóhely és műhely rétegeiből származó pengék hosszainak eltérését szignifikánsnak találja (erre T-próbának nevezve tulajdonképpen Welch-próbát használ, mert a szórásokat –helyesen- nem veszi azonosnak).

Arka 4. lakórétegében talált pengék hossz-szórásainak azonosságát vizsgálja Bartlett-próbával. Elfogadja a szórások azonosságát. A Bartlett-próba alkalmazásának szükséges feltételét, az eloszlás normalitását is vizsgálja.

Helyben, illetve nem helyben található nyersanyag felhasználási arányát vizsgálja a lakó- és munkahelyeken. Látható, hogy az utóbbiakon nagyobb a helyi anyag használati aránya. E tény szignifikanciáját khi^2 -próbával igazolja. Értelemszerűen mind a lakó, mind a munkahelyeken a talált darabokat aszerint sorolja idegen anyagú, vagy helyi anyagú osztályba, hogy a megfelelő darab idegen vagy helyi nyersanyagból készült. Így az osztályok szám 2. Tekintettel arra, hogy a megfelelő khi^2 kifejezés értékére 151 adódott, az ezen értékhez tartozó eloszlásfüggvény értéke , lévén a szabadságfok 2-1=1, sokkal nagyobb, mint 0.999. Emiatt ez nem tekinthető véletlennek, így a munkahelyeken a nem helyben található anyag alkalmazása szignifikánsan (felfelé) tér el a lakóhelyekétől.

Vértes (1965b) módszertana, néhány eredménye, egy kérdés.

Arka lelőhelyénél észleli, hogy a mintaelemek mérete lognormális eloszlást követ. Felismeri, hogy ennek oka lehetne akár véletlen darabolódás is (lásd a fenti megjegyzést a lognormális eloszlással

HU ISSN 1786-271X; urn:nbn:hu-4106 © by the author(s)

kapcsolatban). A lognormális eloszlás alkalmazása ekkoriban kerül be nem-matematikai vagy műszaki területekre egyáltalán.

Sejtett funkcionális eltérések szerinti méretcsoportokra osztva, egy-egy csoporton belül a méreteloszlás normális, ezt grafikusan igazolja a várható érték és szórás becsléssel illesztett normális sűrűségfüggvény és a csoporton belüli hisztogramm együttes ábrázolásával. . Miután e lelőhelyen „sorozatgyártás" valószínűsíthető, a normális eloszlás fellépése elméletileg is várható volt.

Természetesen – és egyes gyártási technológiáknál ez ma is így van- előfordulhat, hogy a sorozat létrehozása után a technológiai egyenetlenségek miatt keletkező „kiugró" darabokat selejtezték. Ekkor a fentiek alapján csonkolt normális eloszlású mintát kellene kapni. Kérdésként merül fel, hogy nagy gyártási számoknál a válogatás e megnyilvánulásával lehet-e találkozni. Ennek észleléséhez azonban a szóban forgónál jóval nagyobb mintanagyság szükséges. Elképzelhető azonban, hogy ilyen darabszámok is előkerülhetnek, de meglátásunk szerint ezt még a neolitban sem tartjuk reálisnak. Talán középkori pénzverő műhelyeknél ez nem kizárható, hiszen ott a tömeg a verés után méréssel ellenőrizhető, s a „kiugró" darabokat olvasztással vissza is vezethették a gyártásba.

Vértes (1967) fő célja a technológiai fejlődés vizsgálata statisztikai módszerekkel.

Felismeri, hogy az eszközök hosszának szórása az időben haladva alig változnak (olduvai (4 réteg) és budai ipar (5 réteg)–Vértesszőllős- adataiból). E látszólagos ellentmondást az oldhatja fel, hogy a megjelenő új eszközök szükségképpen nem annyira kiérleltek.

Bevezeti a rétegenkénti hossz terjedelem (I) és a hossz-szórás (S) I/S hányadosát, és ezen értékre időben monoton növekedő értéket kap, két kivétellel , hozzávéve Combe Grenal (Charentien, Franciaország) 3 rétegét is. Ez- szemléletesen- azt fejezi ki, hogy adott hosszúságúnak szánt eszköz gyártási hosszának ingadozása az időben előrehaladva csökken.

Vizsgálja I/S és az abszolút időskála kapcsolatát. Az (I_1/S_1-I_2/S_2) /(log x_1-log x_2) formula (x idő) értéke a közelebbi múltban vett x_1 és x_2 értékekre szignifikánsan nagyobb, mint a távoli múltban, mely a gyorsuló fejlődést mutatja.

A dolgozat egy lábjegyzetében előrevetíti, hogy Louis Leakey (ekkor még nem elkészült) részletes publikációja Olduvai-ról után részletesebb számításokat fog közölni . Nincs ilyenről ismeretem, ennek oka lehet Vértes László korai, tragikus halála is.

Vértes (1965a) 23. fejezete egy rövid matematikai statisztikai régészeti összefoglaló, Khi^2, F, T-próba régészeti példákkal illusztrált használata. Egy-két képlethibától eltekintve, kiválóan használható, de ismerete nem pótol (nem is tűzi ki feladatának) elemi matematikai statisztikai ismereteket.

Következtetés /zárás

Vértes László kiválóan ismerte és gondosan alkalmazta a matematikai statisztikai módszereket az ősrégészetben. A művekből az is világos, hogy aktív kapcsolatot tartott kora több híres, témával foglalkozó magyar matematikusával. Juvancz Iréneusz és Medgyessy Pál név szerint is említésre kerül, mint egyes, Vértes által végzett számítások ellenőrzője. Sejthető, hogy korának vezető magyar valószínűségszámítási specialistájával, Rényi Alfréddel is kapcsolatot tartott. E sejtést Vértes (1965b) alapozza meg.

Köszönetnyilvánítás

Köszönettel tartozom T. Biró Katalinnak, Dobosi Violának, Gál Erikának, és Bartosiewicz Lászlónak..

Irodalom

PRÉKOPA, András (1962) Valószínűségelmélet műszaki alkalmazásokkal. Műszaki Könyvkiadó Budapest.

RÉNYI, Alfréd (1966) Valószínűségszámítás. Tankönyvkiadó, Budapest.

VÉRTES, László(1964) Das Jungpaleothikum von Arka in Nord-Ungarn.. *Quartär* **15/16** 79-132.

VÉRTES, László (1965a) Az őskőkor és átmeneti kőkor emlékei Magyarországon. Akadémiai Kiadó, Budapest.

VÉRTES, László (1965b) The deposit of silex blades from Boldogkőváralja. *Acta Archaeologica Academiae Scientiarum Hungaricae* **17** 128-136.

VÉRTES, László (1967) Vértes, László Az őskőkor technológiai fejlődési rátái. *MTA II. Oszt. Közleményei* **15** 265-283.

VINCZE, István (1968) Matematikai statisztika ipari alkalmazásokkal. Műszaki Könyvkiadó, Budapest.