

EGYFAJTA VALÓSZÍNŰSÉGI MÓDSZER UTAZÁSI JELLEMZŐK BECSLÉSÉHEZ UTASFELMÉRÉSBŐL

VASS LAJOS

A cikkben egy a szokásos és általánosan használt módszerektől eltérő megközelítésű becslési eljárásról van szó kategorizált változók populációbeli valószínűségeloszlásának, illetve paraméterének számítására. Az eljárás a vizsgált kategória mintabeli gyakoriságának a paraméter függvényében számított valószínűségeloszlásán alapszik. A paraméter becsült értéke pedig a számításhoz felvett paraméter értékek súlyozott átlaga lesz, ahol a súly a mintában realizálódott gyakoriság számított valószínűsége. Lényeges kérdés a mintavételi eljárás modellezése és a valószínűségeloszlás ebből adódó számítása.

A módszer alkalmazása elsősorban egyes utasfelmérési módok esetén az utasfelmérésből történő becslésnél látszik megfontolandónak, a kidolgozását is éppen az utasfelmérés egzaktabb kiértékelése indukálta. Különösen kisebb mintanagyságnál látszik hasznosnak alkalmazása. A cikkben le van írva egy konkrét utasfelmérési módra – járművön történő kikérdezésnél – az eljárás, valamint a számítási mód, valós példán pedig bemutatjuk az alkalmazását a számítások kapott eredményeivel együtt.

1. Bevezetés

Amikor valamilyen jellemző megjelenését, megmutatkozását vizsgáljuk egy populációban, ezt a populációból vett minta alapján tesszük meg. Ha a populáció az emberek bizonyos csoportja, a mintavétel sok esetben a kérdezés, az interjú. A vizsgálat során nem tekinthetünk el attól, hogy a vizsgált jellemző a populációban mennyire változékony; időtől, helytől, stb. mennyire függ. A közlekedéssel, az utazással kapcsolatos jellemzők vizsgálatánál a populáció az utasok valamilyen köre lesz. A jellemzők egy része körükben csak lassan változik, pl. azok, amelyek az utasok utazási szokásaihoz kapcsolódnak, vagy a közlekedési adottságaikhoz kötődnek – a szokások, illetve az adottságok lassan változnak –, de vannak gyorsabban, akár óránként változó jellemzők: utasok összetétele, utazás indoka, a főbb utazási relációk, stb. Amíg az előbbi jellemzőkkel kapcsolatban az utasokat tágabb

időhatárok között kérdezhetjük meg, addig az utóbbi típusú jellemzők vizsgálatánál a mintavétel szokásosan az utasok utazásuk közbeni megkérdezésével történik. Ez utóbbi mintát csak ott és akkor utazó utasok sokaságára lehet vonatkoztatni, vagyis egy adott időintervallumban, vagy egy adott helyen vett minta már nem feltétlen vonatkozik egy-két órával később és máshol utazókra, hiszen az utazás állandóan változó folyamat.

Éppen e miatt egyes közlekedési jellemzők vizsgálata, valamint a rájuk vonatkozó becslési mód eltérhet többé-kevésbé más területek statisztikai vizsgálatától, ez vonatkozik a mintavételre, de a kiértékelésre is. Pl. a kérdezési körülmények, vagy annak az időtartamnak a rövidege, amelyre a mintavételt relevánsnak gondoljuk, nem teszi lehetővé, hogy minden esetben kellő nagyságú mintával számoljunk. Továbbá a kiértékelésnél figyelembe kell venni az utasforgalomra fennálló összefüggéseket is.

E cikk egyfajta becslési módszert tárgyal, amely a sajátosságok miatt a közlekedéssel, az utazással kapcsolatos jellemzők becslésénél látszik jól használhatónak (4. fejezet). A mintavétel egy speciális módja esetében pedig a számítási eljárást ismerteti részletesebben (5., 6., 7. fejezetek). Egy valós példán keresztül pedig a módszer alkalmazását mutatja be a kapott eredményekkel (8. fejezet).

A cikk nem szándékozik foglalkozni részletesebben és matematikai alapossággal az ismertetett becslési mód statisztikai tulajdonságaival, ez még további más jellegű munkát, elemzést és időráfordítást igényelt volna. Inkább kapott hangsúlyt a módszer leírásán túl az alkalmazásának bemutatása egy valós példán keresztül. Ilyen szemszögből egy esettanulmánynak is tekinthető ezen írás.

2. A cikkben használt fogalmak ismertetése

A megértés és az egyértelműség érdekében áttekintjük a használt fogalmakat.

Utazási jellemző: térben, időben, vagy egyéb szempontból (pl. üzemeltető szerint, vagy közlekedési mód alapján) behatárolt személyközlekedés mennyiségét, minőségét, milyenségét összefogóan megadó tulajdonság, ismerv. Amennyiben számértékkel adjuk meg, akkor ezt az értéket az adott jellemző mutatójának nevezzük a specifikált közlekedésre. Ilyen jellemző lehet például egy adott település egy napi közlekedési teljesítménye (utaskilométerben), vagy pl. a délelőtti időszakban az utasok valamilyen szempont szerinti összetétele, százalékos megoszlása.

Az adott közlekedésben résztvevő utasok adják a vizsgált populációt. A populáció egyes egyedeinek (utasoknak) utazási jellemzőit, megkülönböztetésül a közlekedés egészére vonatkozó előbb definiált jellemzőtől, ezen írás keretében ezután *egyéni utazási jellemzőnek* hívjuk.

A közlekedés egészére vonatkozó mutatók egy részét a populáció egyedeinek egyéni utazási jellemzőiből számoljuk ki. Ha nem ismerjük minden egyedre a kérdéses jellemzőjét, akkor mintavétel alapján becsüljük meg a mutató populációra

vonatkozó értékét. Az „utazási jellemző” – szóhasználat előfordul – a fogalmazás gördülékenysége érdekében esetenként az egyéni utazási jellemzőre is. A szövegkörnyezetből kiderül azonban, hogy a két fenti fogalom melyikét kell érteni alatta.

Ezen írás során az egyéni utazási jellemzőket *valószínűségi változóknak* tekintjük, amelyek mindegyike valamilyen saját eloszlást követ a populációban. A valószínűségi változó jellemzésénél a [18] irodalom fogalmait követjük. A valószínűségi változó, az egyéni utazási jellemző milyenségétől függően, felvehet számértéket (metrikus valószínűségi változó), pl. átszállások száma, utazási idő, illetve több nominális értéket, ún. kategóriákat. Ez utóbbi esetben beszélünk kategorizált valószínűségi változókról. Például, ha a valószínűségi változó az **utások státusza**, akkor *dolgozó, tanuló, nyugdíjas, egyéb* értékeket rendelhetünk hozzá, vagy az **utazás indoka** esetén *munkavégzés, tanulás, szórakozás, egyéb* értékeket.

E cikk keretében ez utóbbi esetet tárgyaljuk, vagyis a valószínűségi változó diszkrét, véges számú kategóriaértéket vesz fel. A közlekedés egészét ilyen típusú egyéni utazási jellemző esetén a valószínűségi változó populációbeli eloszlása jellemzi, másként fogalmazva az egyes értékek relatív gyakorisága (százalékos eloszlása). Ez tekinthető ilyen esetekben az utazási jellemző mutatójának.

A valószínűségi változó értékeinek populációbeli relatív gyakoriságát a *populáció paraméterének* nevezzük és $\mathbf{p}$ -vel jelöljük. (A $\mathbf{p}$ vektorként van jelölve, mivel a valószínűségi változó minden lehetséges értékére van egy értéke. Az i -edik komponense p_i az i -edik kategóriára vonatkozó paraméter). Ezen írás a populációban a kategorizált változók paramétereinek becsléséről szól. Ha nincs külön jelölve, akkor paraméter kifejezés alatt valamelyik i -edik vektorkomponensét értjük.

E cikk keretében ezután *alapsokaságnak* nevezzük az utasok azon körét, amelyből a mintát vesszük. A *populációnak* pedig az utasok azon körét, amelyre a becslést végezzük. A populáció lehet az alapsokaság maga, vagy annak részhalmaza. Általában a két fogalmat szinonimaként használják, de mostan tárgyalásunk során célszerűnek látjuk, hogy különbséget tegyünk. Rendszerint a vett mintának csak azon elemeit használjuk fel egy konkrét halmazra vonatkozó becslésnél, amelyek egyúttal ebből a halmazból is származnak. Pl. egy település délelőtti közlekedési mutatóit a település délelőtti utasainak mintájából becsljük, jóllehet a mintavétel egész nap történt. Vagyis az a halmaz, amelyre becsléseket teszünk, megegyezik azzal, amelyből származó mintaelemekből tesszük a becslést.

A közlekedésre vett mintáknál azonban, így a mostani tárgyalásunkban bemutatott esetben is, a két halmaz nem esik feltétlenül egybe. Az alapsokaságból vett minden mintaelem figyelembe van véve a vizsgált populáció paraméterének becslésénél, sőt a számításnál figyelembe vett egyéb feltételek is az alapsokaságra vonatkoznak. Ezért célszerű a két fogalmat ezúttal megkülönböztetni. Például a később ismertetendő EMAH projekt keretében a vonaton levő összes utasra történt az utasszámlálás, de csak a határon átutazókra végzünk becslést.

A rövidség kedvéért a továbbiakban a valószínűségi változó egy értékének (kategóriájának) *mintabeli gyakorisága*, illetve *mintabeli aránya* kifejezés alatt a

változó adott értékének a mintában levő gyakoriságát, illetve relatív gyakoriságát értjük.

3. Utasfelmérés módjai

Számos utazási jellemző az utasok egy részének megkérdezésével kapott ún. *célforgalmi* mintából ismerhető meg. Ilyen felmérések esetében az alapsokaság az utasokat, az utazó közönséget jelenti, illetve a vizsgálattól függően, ezek egy helyben vagy időben behatárolt és konkrétabb halmazát, amelyből a mintavétel történik. Például egy város egy napi utasait, vagy egy megye, egy térség utasait, vagy egy vasúti vonalon utazókat, stb...

A mintavételi eljárást az utazási módhoz alkalmazva, pl. a közforgalmú járművekkel történő (autóbusz, vonat) utazásnál, valamint az adottságokra, a megvalósíthatóságra tekintettel és a gazdaságosság szem előtt tartásával a vizsgálat céljához kell megválasztani. Ennek megtervezése és végrehajtása a felmérni kívánt utazások alapos, részletes ismeretét és nagy gyakorlatot igényel. Autóbusznál, vonatonál általában két módszer használatos az utasok kikérdezésére: *a megállókban történő kikérdezés*, vagy *a járműveken történő kikérdezés*.

Mindkét felmérési mód utazás közbeni csoportos egyszerű mintavétel – amikor is minden szóbjáphető, vagy fontos csoportból veszünk mintát –, amelynek során törekszenek az utasok független és véletlen kiválasztására. Ezek megvalósulását a kiértékelés során tényként is kezeljük. Az utasok kiválasztásának módja azonban nem jelenti azt, hogy a mintában a mintaelemek egymástól függetlenek is!

A korrekt és jól használható felmérésnél a célforgalmi mintavételhez szükséges továbbá az is, hogy kiegészítse utasszámlálás, az ún. *keresztmetszeti felmérés*. Ez a megállóhelyeken fel- és leszálló utasok megszámolását jelenti. A populáció, illetve az alapsokaság nagyságát kapjuk meg ez által. Egy adott felmérésnél ezeket az utasszámokat nem tekintjük véletlen értékeknek.

Az utasfelmérés mindegyikére jellemző, különösen, ha időben, térben kiterjedt utasforgalomról van szó, hogy költségessége miatt nem ismétlik meg hasonló körülmények között, ezért általában nincs több minta, csak egy. Ezért egy minta alapján kell a hibát is megmondani, ha egyáltalán megtudjuk. Általában több ezres utasforgalom esetén mintegy 10-15%-os mintavételi arány a szokásos elvárás (utalunk itt a KTI-ben levő számos utasközlekedéssel foglalkozó tanulmányra). Ebben az esetben a teljes alapsokaságra – feltéve a normális eloszlás közelítést a kategóriák mintabeli relatív gyakoriságára – a paraméter hibája nagy biztonsággal ($\approx 90\%$) 0,05 érték alatt marad. Ahhoz, hogy egyes fontosabb utascsoportokra kellően nagy mintadarabszám jusson, a kikérdezés során minél nagyobb mintavételi arány elérésére törekszenek. Azonban az, hogy az egyes vizsgált részpopulációkba hány darab mintaelem kerül, az majd a feldolgozás során derül ki. A kisebb részpopulációkra már csak nagyobb pontatlansággal számolhatunk. Ezért ezekre nem lehet hibát előre megfogalmazni.

4. Paraméter felmérésből történő becslése

4.1. Jelenleg szokásos eljárás

Az egyéni utazási jellemző, mint valószínűségi változó, eloszlását megadó p paraméter becslési értéke szokásosan a jellemző egyes kategóriáinak a vizsgált populáció célforgalmi mintájában levő aránya. Ezt az arányt alkalmazzuk a populációnál, és így egy kategória populációban előforduló számosságát a populáció nagyságának ezen aránnyal való megszorzásával becsüljük meg. Ezt gyakran a minta „felszorzásának” is nevezik.

Az arány alkalmazása praktikus egyszerűsége és a gyakorlati esetek nagy részében elfogadható pontossága miatt. Vannak olyan vizsgálatok azonban, ahol az arány alkalmazása kétségeket vet fel.

Amennyiben a kikérdezéses minta jól visszatükrözi a populációt, akkor a mintából jól becsülhető az utazási jellemző populációbeli eloszlása. Például így van ez akkor, eloszlástól függetlenül, ha a mintanagyság közelíti a populáció számosságát, vagy a minta kellően nagy egy adott pontossághoz, például a Bernstein tétele értelmében.

Amennyiben egy kiválasztott kategóriát nézünk, akkor tekinthetjük úgy, hogy a populáció egyedei csak kétféle értéket vesznek fel (az adott kategóriát, illetve nem azt), vagyis a valószínűségi változó Bernoulli-eloszlást követ. Ilyen esetekben a célforgalmi felmérés mintavételi eljárása alapján a kategória mintában levő gyakorisága esetenként binomiális vagy hipergeometrikus eloszlású lehet. A paraméterre a relatív gyakoriság ilyen esetekben jó becslést ad: torzítatlan, konzisztens és a becslés hibáját is ismerjük, $1/\sqrt{n}$ -nel (ahol n a minta elemszáma) arányos. Ezen eloszlásokra a maximum likelihood módszerrel is a relatív gyakoriságot kapjuk torzítatlan és konzisztens becslésként, hipergeometrikus eloszlásnál aszimptotikusan. [2], [12]. E fenti eloszlások általában csak egyszerűbb felmérésnél és mintavételnél jöhetnek létre. „Tiszta” eloszlással általában nem számolhatunk a gyakorlatban megvalósuló mintavételi eljárások során.

Amikor egész napi, vagy nagyobb területre, pl. egész városra vonatkozó utazási mintát nézünk, akkor egyes populációk (pl. ha egyúttal az maga az alapsokaság), valamint a minta is elég nagyok lehetnek, így a nagy minta esetére alkalmazhatjuk a központi határeloszlás tételét a mintaelemek eloszlásának széles körére, ami szerint a relatív gyakoriság jól közelíthető normális eloszlással. Nem csak független és azonos eloszlású mintaelemek esetében alkalmazhatjuk, de különböző eloszlásúakra, sőt nem függetlenekre is [16], [17]. Ilyen esetben is a mintabeli arány a populációra jó becslést ad, és ismerjük a hibáját is. Utalunk itt a 3. fejezet végére. Eközben azonban eltekintünk a mintavétel eltérő tér- és időbeli történésétől. Ahogy említettük a **Bevezető**ben, az egész minta felhasználása a közlekedés egészére ad meg egy átlagos értéket az utazási jellemző mutatójára.

Egyes elemzéseknél azonban érdekelhet minket, hogy az utazási jellemző időben hogyan változik, pl. hogyan alakul egy mutató óránkénti értéke, vagy kisebb

térségben (pl. egy város valamelyik körzetében) mik lesznek a jellemzők értékei. Ekkor az egész minta valamilyen kiválasztott részét nézzük csak, amelynél a minta nagysága, vagy a mintavételi arány is kicsi, és amelyben nem ismerjük a kategória eloszlását, és nem alkalmazhatjuk a központi határeloszlás tételét sem kellő pontossággal. ([16]-ban a normális eloszlás alkalmazásának elfogadhatóságára adnak meg összefüggést.) Ezért kisebb mintáknál a mintabeli arány eloszlásáról nem tudunk semmit, így a paraméter becslésére ezen érték felhasználása kétséges megbízhatóságú, hibáját sem tudjuk megmondani. S mivel csak egy minta van – mint már említettük, a felmérést általában nem ismétlik meg –, még egy másik méréssel sem tudjuk összevetni a kapott becslést.

Gyengébb minta esetén alkalmazhatunk önkényes megfontolásokat is, amelyek gyakorlati szempontból elfogadhatók és általában megfelelő eredményeket adnak, de matematikai-statisztikai szempontból nem értelmezhetők. Ezért elméletileg jobban alátámasztható más becslési módot dolgoztunk ki. Ez a matematikai-statisztika és a valószínűségszámítás módszereinek alkalmazásán alapul, lehet vele becsülni a standard hibát is, és emellett figyelembe veszi az utasforgalom összefüggéseit is.

4.2. Egy más megközelítésű eljárás

Alapvetően az ebben a pontban ismertetendő eljárás a cikk tárgya. Az eljárás lényege, mint általában a becslések esetében: a paraméter értéke a kategória mintabeli gyakoriságának valószínűségeloszlása alapján lesz becsülve. Ez az eloszlás függhet a mintavételül szolgáló alapsokaság és a populáció nagyságától, a paraméter értékétől, a célforgalmi kikérdezés módjától, mikéntjétől, valamint a mintavételi eredmény realizációjától (pl. mikor, milyen mintaelemeket kaptunk). Az általunk használt megközelítésnél a becslési eljárásnak két fő része van: először a vizsgált kategóriára a mintabeli gyakoriság valószínűségeloszlásának meghatározása a paraméter függvényében, azután a valószínűségeloszlásból a p paraméterre becsülő statisztika készítése. A becsülő statisztika egyfajta súlyozott átlag lesz. Lásd alább. Ebből a statisztikából a becslés hibáját is meghatározhatjuk.

Nevezzük *jóesetnek* az általánosság és a rövidség kedvéért, ha egy utas az alapsokaságban a vizsgált kategóriájú utazást bonyolítja le. A jóeset előfordulhat a mintában is, oda azonban véletlenül kerül be. Ha bevezetjük az indikátorváltozót arra az eseményre, hogy egy adott mintaelem jóesetként szerepel a mintában, és 1 értéket rendelünk hozzá, és 0-át az olyan mintaelemhez, amely nem egy jóesetet jelenít meg, akkor az indikátorváltozó értékeinek összege a mintára egy statisztika, és *véletlen változó*, amely a jóesetek mintában levő számát adja meg.

Legyen ez a véletlen változó ξ , és P legyen annak valószínűsége, hogy $\xi = k$ érték adódik n utas kikérdezéséből (n nagyságú a minta), és ez a populáció p paraméterének, a populáció M , valamint a mintavételül szolgáló sokaság N nagyságának (utasok száma) függvénye. Vagyis:

$$P\{\xi = k; p, n, M, N\} = P_k(p; n, M, N).$$

Általában és leggyakrabban M -et és N -et azonosnak vesszük a mostani írás kivételével. Ez a függvény a vizsgált kategóriához tartozó utazások mintabeli gyakoriságának valószínűségeloszlása.

A P függvény formája akár megadható zárt alakban, akár nem, függ a mintavételi eljárástól és a mintavételi történésektől. Adott n , M , N esetén $\xi = K$ előfordulás valószínűsége, a $P_K(p; n, M, N)$ függvény, ahol K a mintavétellel kapott jóesetek száma, csak a populáció paraméterétől függ, és e függvény alapján a paramétert megbecsülhetjük.

$P_K(p; n, M, N)$ függvényt ábrázolhatjuk adott K esetén p függvényében. A p paraméter becslésének egyik lehetősége az a p_{max} érték, amelynél a függvény a maximális értékét veszi fel a fenti feltételek esetén p szerint, vagyis

$$P_K(p_{max}, n, M, N) = \max P_K(p; n, M, N).$$

A p_{max} érték a paraméter maximum likelihood becslése. Egy vizsgált kategóriának mintabeli aránya lehet a p paraméter becsült értéke a maximum likelihood módszer alkalmazásával bizonyos eloszlásoknál, ahogy már szó is volt róla.

Az általunk alkalmazott megközelítésben a p paraméter becslésének eljárása általánosságban a következő lépések szerint történik.

1. A mintavételi mód modellezése.
2. A P függvény (valószínűségeloszlás) számítási eljárásának meghatározása.
3. Paraméter függvényében a P függvény számítása.
 - a) A p paraméterre egy lehetséges p_r értéket felvéve, számítjuk a ξ valószínűségi változó eloszlását a mintában, és a bekövetkezett K előfordulás valószínűségét.
 - b) A valószínűségi értéket a paraméterre felvett p_r értékkel együtt elmentjük.
 - c) Vesszünk egy következő lehetséges értéket a paraméterre, és a 3. a) ponttól folytatjuk
4. Az elmentett értékpárok alapján felvesszük a valószínűség-paraméter függvényt, és a paraméter becsült értéke a p_r értékeknek a számított valószínűséggel súlyozott átlaga lesz.

A 3. pontból látható, hogy a számítást numerikusan végezzük, hiszen nem várható, hogy valamilyen analitikus formájú eloszlásfüggvényt kapjunk.

A numerikus eljárás röviden

A számítást az egyszerűség kedvéért diszkrét p értékekre végezzük el. A p paraméter lehetséges értékét tartalmazó $[0, 1]$ intervallumot R egyenlő diszjunkt

intervallumokra osztjuk fel. Az egyes intervallumok kiválasztott pontjai legyenek a $p_1, p_2, \dots, p_r, \dots, p_R$ értékek, ezek lesznek a paraméterre felveendő lehetséges értékek, amelyekre aztán a 3-as pont végrehajtott.

Az R nagysága a számítás pontosságát meghatározza, értékét a gyakorlati esetben elfogadható pontossághoz lehet megválasztani. A számításainkban a $[0, 1]$ intervallumot 100 egyenlő részre osztottuk fel, hogy a számítás elég pontos is legyen, időigénye pedig elfogadható.

Megjegyzés: ha tudjuk, hogy a paraméter milyen intervallumba esik, akkor elegendő csak azt az intervallumot vizsgálni, és ott venni fel szükséges számú p_r értéket.

A 4. pontnál a becslő statisztika a valószínűséggel súlyozott átlag lesz:

$$p_{becs} = p_{sátl} = \sum p_r \cdot P_K(p_r; n, M, N),$$

ahol az összegzés $r = 1, 2, \dots, R$ -re történik, továbbá fenn kell álljon:

$$\sum P_K(p_r; n, M, N) = 1.$$

4.3. Az eljárás összevetése más módszerekkel

A fentebb leírt eljárás egyik fő lépése, a mintabeli gyakoriság valószínűségének számítása, magától értetődő eljárás, ez a maximum likelihood becsléshez is szükséges lépés. A másik fő lépés a p becslt értékének valószínűséggel súlyozott átlaggal történő számítása. Szűkebb internetes irodalmi kutatás során nem volt fellelhető irodalmi példa ilyen statisztikára. Természetesen nem zárható ki, hogy ilyen becslő statisztika alkalmazásának már volt akár több előzménye is. Az e statisztikával történő becslésnél számíthatjuk a becslt érték várható értékét, illetve a becslés standard hibáját, amennyiben adott mintavételi körülmények mellett a mintában előfordulható jóesetek számát tekintjük valószínűségi változónak. E cikk keretében nem tárgyaljuk a várható érték és standard hiba számítását. Ez a becslési mód a maximumra nem szimmetrikus eloszlásokra nem a maximum érték szerinti becslést adja. A súlyozott átlaggal való becslés statisztikai tulajdonságait sem tárgyaljuk, de egy-két szempontból összevetjük más módszerekkel.

A *maximum likelihood* becsléssel szemben a súlyozott átlag jobban figyelembe vesz két szempontot. Egyrészt az eloszlásfüggvény olyan tulajdonságát, hogy a maximum értéken kívüli más p értékek esetén is viszonylag nagy valószínűséggel előfordulhat a mintában a kapott K érték, tehát egy másik p érték is jó eséllyel lehet a populáció paramétere. Valamint, hogy elkerüljük azt az esetet, mikor a maximum a paraméter 0 értékénél van, ugyanis $p = 0$ eset a közlekedésben ritkán fordul elő. Gyakrabban viszont az, hogy nincs a mintában mintaelem valamely kategóriára.

A paraméter becslésére, becslés hibájának vagy egyéb becslési jellemzőnek számítására használják a kifejezetten számítógépes alapú *bootstrap* eljárást. Ez a

módszer azonos eloszlású és független elemeket tételez fel a mintában [17]. Az utasfelmérésre jellemző mintavételeknél ezzel általában nem számolhatunk, továbbá az eloszlást egyrészt nem lehet analitikusan megmondani, másrészt az a nem ismert paraméter függvénye is. Ezért – ha szóba jöhet – csak a paraméteres bootstrap jöhet számításba. Esetünkben, leegyszerűsítve, a mintában jóesetek és nem jóesetek fordulnak elő, vagyis kétféle értékű adatsorunk van. (A megállónban történő mintavételnél általában többértékű adatsor jön számításba.) A mintából valamilyen módszerrel (így akár az általunk kidolgozottal) megbecsülhetünk egy közelítő értéket a paraméterre, és a kapott paraméterérték alapján generálhatunk bootstrap mintákat az adatsorunkból a minta minden egyes elemének kiválasztási valószínűsége alapján. A mintavételi modellünk alapján számolt valószínűségi értékeket itt fel kell használni. Lehetséges, hogy a mintaelemek függősége miatt szukcesszív módon kell eljárni. A kapott mintára pedig valamilyen módszerrel számítjuk ki a paraméter újabb becsült értékét. A 6. fejezetben tárgyalt esetben is eljárhatnánk így. A számítás elvégezhető, de túlságosan időigényes, és a gyakorlati pontosság sem igényli.

Itt megemlíthető a bootstrap módszer egyik alkalmazási lehetősége. Elébe menve a 8. fejezet 8.4.3. pontban tárgyalt *érzékenység vizsgálatnak*, a bemenő adatokban történő változásnak valamekkora hatása van a végeredmény változásaira (ez az ún. érzékenység), különösen kis minták esetén. E kiadvány pár éve foglalkozott egy összefoglaló cikkben [10] a paraméterek becslését általában befolyásoló tényezőkről és a különböző módszerekről. Több irodalmi utalás van benne említve a szimulációs, nevezetesen a bootstrap technika alkalmazására, olyan esetekben, amikor az eloszlás nem formulázható és a mintaelemszám kicsi, amint ez esetünkben is fennáll.

Egyes esetekben a mintában kapott gyakoriság (K) az adott mintavétel jellemző értékei alapján egy túlzottan kicsi, vagy éppen nagy valószínűségű esemény bekövetkezésének eredménye lesz. Ha ezt használjuk fel a becslésnél, akkor nagyobb hibával számolhatunk. Használhatunk *szimulációs módszert* arra, hogy egy valószínűbb jóeset számot kapjunk a mintában. A szimulációról és alkalmazásairól számos irodalom van. Jó áttekintést ad ezekről [8], [9]. Az alkalmazásához az eloszlás ismeretére szükség van, hogy véletlen változót generáljunk. Ahogy a paraméteres bootstrap alkalmazásához, úgy a szimulációnál is a paraméter értékének előzetes becslése szükséges. A becsült paraméterértéknél kapott eloszlás alapján jóeseteket generálhatunk. A mintára ezen számok mellett aztán a paraméter számítására az általunk alkalmazott eljárást elvégezhetjük. Ez sok számítást igényelhet. A számítást kevesebbszer kell elvégezni, ha a generált jóeset számok átlagát vesszük. Ezen átlag-gyakoriságot tekinthetjük a mintabeli pontosabb gyakoriságnak, és csak erre a gyakorisági értékre végezzük el a számítást és kaphatunk egy újabb becslést a paraméterre. De az átlagtól eltérő gyakoriságra megismételve a számítást, akár standard hibát vagy konfidencia intervallumot is számíthatunk. Ebben az eset-

ben a szimulációs módszert nem egy bonyolult eloszlás számítására, vagy valamely becslési jellemző számítására használjuk, hanem csak új mintaelemek generálására.

A tárgyalt módszer alkalmazását a 6. fejezetben – egy adott konkrét mintavételi eljárás esetében (járművön történő kikérdezésnél) – bemutatjuk a megálló közötti utasmozgások számítására. Felmerülhet erre más lehetséges számítási eljárás alkalmazása is. Ennek néhány vonatkozását a 6. fejezet alatt tárgyaljuk.

A fenti 4.2. részben leírt eljárás a megállóban történő kikérdezés esetén tárgyalva lett a Közlekedéstudományi Intézet (KTI) 2011–2012-es évkönyvében [21]. A járművön történő kikérdezésnél történő alkalmazása és a konkrét számítási eljárás e cikk tárgyát képezik.

5. Mintavételi modell a járművön történő kikérdezésnél

A járművön történő mintavételkor a járművön levő utasokat kérdezik meg, és a már kikérdezett utast a felmérés során, a megkérdezés napján, nem kérdezik meg többé. A járművön levő utasok száma állandóan változik, de nem folyamatosan, hanem lépcsőzetesen, két megálló közötti szakaszon állandó az utasszám. A kapott minta tartalmazza, hogy melyik felszálló megállóból melyik leszállóban szállnak le az utasok. A kikérdezést kiegészíti még a járműre történő fel- és leszálló utasok számlálása megállónként.

Mivel minden jármű (pontosabb beszélni a közforgalmú közlekedésnél *járatról* vagy vonatról, amely meghatározott időben meghatározott útvonalon közlekedő járművet jelent) időben máskor közlekedik, akár eltérő útvonalon, részben vagy egészben más megállókat érintve, ezért az utasok összetétele, utazási attitűdjük járatonként változhat, ezáltal az egyéni utazási jellemzők eloszlása is. Ezért minden egyes járat utasaira külön becslés jön szóba.

Alapsokaságnak tehát a járat utasainak összességét tekintjük, közülük történik a mintavétel. Azonban paramétereket erre az alapsokaságra vonatkoztatni a feladathoz, az elemzéshez nem minden esetben megfelelő, mivel egy járat utasai általában több csoportba sorolhatók utazási jellemzőik (mutatók) tekintetében – amelyek esetenként jelentősen különbözhetnek –, és ezért a járatot nem lehet egy homogén egységnek tekinteni, valamint sok esetben ezen különböző csoportok jellemzői érdekelhetnek minket. Az utasoknak olyan körét kell definiálni, amelyre az utazási jellemzők viszonylag egyértelműen erre a körre vonatkoztathatók. Továbbá kell, hogy legyen ismeretünk a csoport nagyságáról. A mintát is majd erre a halmazra kell „felszorozni”, vagyis ez lesz a populáció. Ez a populáció nem esik feltétlenül egybe az alapsokasággal.

Egy járatnál minden egyes megállóhoz hozzárendelünk egy populációt, éspe dig a megállóban felszállt vagy leszállt utasok halmazát. Azt mondhatjuk, hogy egy járatnál egy adott megállóban felszállt utasok az utazási jellemzők olyan mutatóit határozzák meg, amelyek az adott megállóra jellemzőek, egyben különböz-

hetnek más megállók utasai által meghatározott mutatóktól. Tehát megállónként más és más lehet az utasok összetétele, az utazás indokának megoszlása, stb... Ehhez a halmazhoz hozzárendelhetők az egyéni utazási jellemzők eloszlását leíró paraméterek. Ugyanez igaz a leszálló megállókra is. (Itt megjegyezzük, hogy a megálló összes utasforgalma ismert. De nem lehet alkalmas populáció pl. a megállóban felszállt dolgozók összessége, mert azok száma már nem ismert.)

A következő részek tárgyalásánál az egyéni utazási jellemző az **utazási reláció** lesz. Vagyis az a jellemző, hogy adott megállóban felszállt utas melyik megállóba utazik. Ehhez az utazási jellemzőhöz, a 2. pont értelmében, hozzárendelt valószínűségi változó értékei (kategóriák) pedig lehetnek pl. az egyes megállók sorszáma a járaton. A valószínűségi változó csak a hozzátartozó populációra vonatkozik. A valószínűségi változó eloszlása adja meg, hogy milyen arányban oszlanak el egy felszálló megálló utasai a különböző leszálló megállók között, vagy más oldalról, a leszálló megálló szemszögéből, a felszálló megállók között. Ezt adják meg az adott populációra vonatkozó paraméterek. Minden megállóhoz, illetve a populációjához hozzárendelünk valószínűségi változót, és minden egyes változóhoz annyi elemű paramétervektor tartozik, ahány megállóban leszállhatnak, vagy felszállhatnak az utasok.

A megállók közötti szakaszon több megállóból jövő (vagy más megközelítés esetén több megállóban leszálló) utas tartózkodik, tehát egyszerre több megálló populációja létezik együtt, tehát egyszerre több és különböző paraméterű populáció van jelen. Amennyiben megállók közötti mintabeli forgalom (mintabeli gyakoriság) valószínűségeloszlásának számításánál egyszerre kívánnánk kezelni a megállókat, akkor sokdimenziós problémával kellene szembenézni. Emellett a valószínűségeloszlásokat nem lehet analitikus formában megadni, de még numerikus kezelése is nagymértékben bonyolultnak látszik.

Ezért a problémát leegyszerűsítettük. Egyrészt egyszerre csak egy populációt vizsgálunk. Másrészt az egyes megállókban a leszállók számát a mintában nem tekintjük valószínűségi változónak, és értéke a megálló mintában realizálódott leszálló utasszám lesz. Ez azt jelenti, hogy az utasgyakoriság eloszlásának számításánál az egyes megállónál leszálló jóesetek száma nem haladhatja meg a megállóra vonatkozó ezen értéket. E feltétel mellett minden lehetséges eset figyelembe van véve, és a nem megvalósulható esetek (pl. a mintában egy szakaszon (két megálló között) mintavételkor nem kaphatunk annyi jóesetet, amellyel már meghaladnánk a vizsgált megállóban a leszálló utasszámot) ki vannak zárva. Ez a feladat könnyebben megoldható.

5.1. A leegyszerűsített modell

A leírt utas-kikérdezési módszer egy visszatevés nélküli mintavételhez hasonlít két megálló között. A ξ valószínűségi változóhoz hasonlóan bevezetjük két megálló közötti szakaszra, legyen ez az i -edik szakasz, a ξ_i valószínűségi változót, amely az i -edik szakaszon történt mintavételnél kapott jóesetek számát adja meg. A beve-

zetett valószínűségi változók egy kiválasztott populációra (megállóra) értendők. Ez a változó ún. hipergeometrikus eloszlást fog követni az alábbi forma szerint.

$$H(\xi_i = k; N, M, n) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

ahol két megálló közötti i -edik szakaszon

N az utasszám,

M az utasok között a jóesetek száma,

n a mintavétel száma,

k a szakaszon vett mintában levő jóesetek száma.

A következő szakaszon, vagyis a következő két megálló között történő mintavétel esetén is ugyanez az eloszlásfüggvény, de argumentumai már más értéket vesznek fel, és függenek az előző mintavétel kimenetelétől.

Tehát a mintavétel két megálló közti szakaszon a teljes alapsokaságból, vagy részhalmazából (az addig felszállt utasokból) történik, de a becslést csak egyik konkrét részhalmazára végezzük el: amennyiben a felszálló megállót vizsgáljuk, akkor az adott felszálló, ha a leszálló megállót, akkor az adott leszálló megálló utasaira. A járat utasaira (fel- és leszállók számára) vannak utasszámlálási adatok, de különbséget nem tudunk tenni abban a dologban, hogy pl. egy megállóban leszállók között mennyi a különböző megállókból jövő utasok száma, vagyis melyik populációból valók. Ezért bizonyos számítások az összes, megállónként nem megkülönböztetett utasra fognak vonatkozni. Az összes utas alatt az alapsokaságnak a vizsgált leszálló megállóig felszállt utasok halmazát értjük természetesen.

A jóeset az lesz, ha a vizsgált megállóban felszállt utas a vizsgált utazási jellemzőnek, pontosabban valamely kategóriájának megfelelő utazást bonyolít le. Továbbá, mivel csak egy adott megállóra vonatkozó populációt nézünk, a más megállók utasait úgy vesszük, mint ami nem *jóeset*. Így, amennyiben az utazási jellemző a reláció, a felszálló megálló j , a leszálló ℓ , akkor csak a (j, ℓ) reláció lesz jóeset. Ha a leszálló utasokat vizsgáljuk, akkor is így értelmezzük a jóesetet.

6. A megállók közötti utasszám számítása a populációra

A közlekedés elemzése szempontjából a lényeg: mekkora a forgalom nagysága (utasok száma) megállók között, vagy települések között, stb... Ez gyakorlatilag megfelel annak az utazási jellemzőnek, hogy adott felszállóhelyről hányan utaznak más megállókba, mondjuk egy j megállóban felszállt utasokból a követő megállókban hányan szállnak le. (De lehet más utazási jellemzőket is vizsgálni, például hány dolgozó utazik a járművön az adott megállóból, stb...) E számításhoz az **utazási reláció** egyéni utazási jellemzőnek (valószínűségi változónak) eloszlását

használjuk fel, amelynek értékei (kategóriái) lehetnek például, hogy a j megállót követő megállók közül hányadiknál, elsőnél, másodiknál stb.-nél száll-e le az utas. Ezeket a vizsgálatokat el lehet végezni minden egyes megállóra $j = 1$ -től, az utolsó kivételével. Minden egyes két megálló közötti forgalom nagyságát e valószínűségi változó populációbeli eloszlása tehát paramétere segítségével kaphatjuk meg, megszorozva vele a populáció számosságát. A paramétereket azonban nem ismerjük.

A tárgyalás egyszerűsítése érdekében egy vizsgált megállóban *felszállt* utasok képezzék a populációt. (A másik lehetőség lenne, hogy a vizsgált megálló *leszálló* utasai jelentik a populációt. A számítási elv ugyanaz marad ez utóbbi esetben is, csak egyes konkrét számítási megoldások különböznek a két esetben.)

6.1. A megállók közötti utasszám valószínűségeloszlásának számítása a mintában

Az **utazási reláció** valószínűségi változó esetén a 4.2.-ben leírtak szerint a paraméterek kiszámításához a megállók között utazók mintabeli számának valószínűségét kell először meghatározni. A j -edik felszálló megálló esetében a j -edik és a követő megállók közötti utasszámot. A számítási eljárást most egy kategóriára végezzük el, arra, hogy az ℓ -edik megállóban száll le az utas. A megbecsülendő érték, hogy itt a vizsgált megállóban felszállt utasokból hányan szállnak le milyen valószínűséggel. Jóeset most esetünkben, ha egy utas, amelyik a j -edik megállóban felszállt, az ℓ -adik megállóban száll le.

Tehát a felszállók között $M_{j\ell} = p_{j\ell} \cdot F_j$ lesz az adott jellemzővel bíró utasok (jóesetek) száma, ahol

$M_{j\ell}$ az ℓ -edik megállóhelyen leszállók száma a j -edik megállóból,

F_j a j -edik megállóhelyen a felszálló utasok száma,

$p_{j\ell}$ az ℓ -edik megállóra jellemző paraméter a j -edik megállóban felszállók körében.

Annak valószínűségét, hogy a $j \leq i$ -edik és az $i + 1 \leq \ell$ -edik megálló közötti i -edik szakaszon vett mintában a j megállóban felszállt utasok közül k darab olyan utas fordul elő, amelyik az ℓ -edik megállóban leszáll, az alábbiak szerint írhatjuk fel, amennyiben a megelőző szakaszokon eddig összesen z_i darab jóeset fordult elő.

$$P(\xi_i = k) = P_k(i) = H(k; N_i, M_i, n_i | C_{j\ell}) = H(k; N_i, M_{j\ell} - z_i, n_i | C_{j\ell}), \quad (1)$$

ahol

i az argumentumban az összes szükséges paraméter i -edik szakaszbeli értékére utal,

N_i az adott szakaszon figyelembe vehető utasok száma,

M_i az adott szakaszon figyelembe vehető jóesetek száma,

n_i az adott szakaszon a mintavételek száma,

$C_{j\ell}$ olyan feltételek, amelyeknek figyelembevételével kell a valószínűséget kiszámítani.

Nem részletezve, ezek arra vonatkoznak, hogy a k szám nem lehet tetszőleges nagy egy szakaszon, korlátot szab az ℓ -edik megállóban leszálló utasok száma, valamint megállónkénti leszállók száma a mintában, szakaszonként vett mintaelemek száma, továbbá a vizsgált szakaszig az összes vett jóeset száma (z_i értéke) is.

A figyelembe vehető utasszám:

$N_i = (\text{az utasok száma az } i\text{-edik szakaszon}) - (\text{az előző szakaszokon megkérdezett és az } i\text{-edik szakaszon még fennlevő utasok}).$

A figyelembe vehető jóesetek száma:

$M_i = M_{j\ell} - (\text{az } i\text{-edik szakaszt megelőző szakaszokon megkérdezett utasok között a jóesetet jelentő összes utas száma, vagyis } z_i).$

Ha az i -edik szakaszig z_i számú jóeset fordul elő a mintában, akkor a $P_k(i)$ valószínűséget felfoghatjuk feltételes valószínűségnek is, ezért jelölhetjük $P_k(i|z_i)$ -vel a függvényt. Így az előző egyenletet (2) szerint is felírhatjuk.

$$P_k(i|z_i) = H \{k; N_i, M_{j\ell} - z_i, n_i | C_{j\ell}\}. \quad (2)$$

Annak valószínűségét, hogy egyáltalán az i -edik szakasszal bezárólag, azaz az i szakaszon kapott jóeset számmal együtt összesen w jóeset forduljon elő, a (3) alapján számíthatjuk ki, mivel a z_i -nek több értéke lehet véletlen módon. S_i jelöli az i -edik szakasz utáni jóesetek számának valószínűségeloszlását:

$$S_i(w) = \sum_{z_i=0}^{z_{\max}} P_{w-z_i}(i|z_i) * S_{(i-1)}(z_i), \quad (3)$$

ahol az összegzés z_i -re a megelőző szakaszokon maximálisan elérhető lehetséges értékekig, $z_{\max}$ -ig történik.

A w érték pedig nem lehet több, mint $z_{\max} + k_{\max}$. A $k_{\max}$ érték az i -edik szakaszon kiválasztható maximális jóesetek száma, amelyet az i -edik szakaszon vett minták száma és a z_i érték határoz meg. A (3) összefüggés a teljes valószínűség tétele alapján lett felírva, mert a különböző z_i értékek előfordulása egymást kizáró események, és valószínűségeik összege 1. Ha $w - z_i < 0$, akkor a feltételes valószínűség értelemszerűen 0 lesz.

A (3) összefüggés azt mutatja, hogy az $S_i(w)$ valószínűséget szukcesszíve számíthatjuk az előző szakaszokból és az adott szakaszon történt mintavételből. Vagyis az előző szakaszokon kapott összes jóeset valószínűségéből és az aktuális szakaszon kapott jóesetek száma valószínűségéből lehet számolni az aktuális szakasz utáni jóesetek számára vonatkozó eloszlást. Felhasználva, hogy az első szakaszra $z_1 = 0$ lesz és $S_0(0) = 1$, hiszen az első szakasz előtt nem volt mintavétel. Minden szakasz után megkapjuk az addig lehetséges jóesetek számát, valamint a hozzátartozó való-

színűséget. A kapott jóesetek számával aztán csökkentjük a következő szakaszon figyelembe vehető jóesetek számát $(M_{j\ell} - z_i)$ -t, és folytatjuk a következő $(i + 1)$ szakaszon történő mintavétel valószínűségének számításával. Ezt a számítást elvégezzük minden két szomszédos megálló közötti szakaszra a j -edik megállótól az ℓ -edikig, és így megkapjuk az adott utazási jellemzőre, vagyis a j megállóban felszállt és az ℓ -nél leszálló utasok számának eloszlását a mintában adott $p_{j\ell}$ mellett. A paraméter ezen értékénél így megkapjuk a $\xi = \sum \xi_i \quad i = j, \quad j+1, \quad j+2, \dots, (\ell-1)$ valószínűségi változó eloszlását.

A gyakoriság valószínűség eloszlásának fenti számítását numerikusan végezzük a 4.2.-ben a 3-as pontban leírtak szerint a $p_{j\ell}$ paraméterre felvett minden $p_r, r = 1, 2, \dots, R$ értékre, és így a mintában az ℓ -nél leszállók számához tartozó valószínűségeket is ismerjük minden p_r -re, és abból a p_r -ek súlyozott átlagát tudjuk számítani. Így kaphatunk becslést a paraméter értékére.

Elvégezhetjük a számítást minden szóbajöhető (j, ℓ) megállópárra, rögzített j mellett, ahol $j < \ell$, továbbá minden j -re. Mivel minden mintaelemről tudjuk, hogy melyik szakaszon vették, és melyik megállóban szállt le, továbbá ismerjük a megállónkénti felszálló és leszálló utasok számát, a fenti számítást elvégezhetjük. Nem szükséges minden megállópárra elvégezni, ha csak egy megállóban felszállt utasok mozgása érdekel, vagy csak adott megállóban leszállt utasok száma érdekel bennünket, akkor csak j -ben felszállt, vagy csak ℓ -ben leszálló utasokat kell nézni. Ilyen eset lehet például, ha a megálló egy átszállóhely.

A (3) összefüggés azt mutatja, hogy mivel minden, az adott szakasznál lehetséges z értékre történik az összegzés, az eloszlás számításánál az összes megvalósulható eseményt figyelembe vesszük, és nem vagyunk tekintettel arra, hogy hogyan realizálódott az egyes szakaszokon vett mintavételek során a mintában végül jelenlevő jóesetek száma. Vagyis a konkrét realizáció figyelembe vétele nélkül (a priori) számítjuk az eloszlást. Ez az 5.1.-ben leírtaknak megfelel, vagyis minden lehetséges eset alapján számítjuk az eloszlást.

A számítást kiegészíti, túl az $C_{j\ell}$ feltételek figyelembevételén, két alapvető feltétel számítása: az ℓ -edik megállóban a *lehetséges maximális leszállók* számának, illetve a *legalább szükséges leszállók* számának számítása. Ezeket a számokat a keresztmetszeti felmérésből kapjuk és a konkrét utasforgalom határozza meg. A lehetséges maximális leszállószám azt adja meg, hogy a j megállóban felszállt utasokból mennyi szállhat le legfeljebb az ℓ -edik megállóban. (Ha közben kevés a felszálló utas, akkor a j -edik megállóból más közbelső megállókba is kell kerülni utas, ezért korlátozva van j -edik megállóból az ℓ -edik megállóban leszállható utasok száma.) A legalább szükséges leszállók száma pedig azt adja meg, hogy a j -edik megálló utasai közül ennyinek biztos le kell szállnia ℓ -ben. (Ha ℓ -ig kevés a felszálló utas más megállókban az ℓ -nél leszállókhoz képest, akkor j -ből is kell, hogy valamennyi leszálljon.) Így az eloszlás számításánál, illetve a becslésnél a számlált utasforgalom alakulása is figyelembe van véve. Ha nincs minta véve valamely

(j, ℓ) megálló pár között, akkor a fenti két szám átlagát tekinthetjük a paraméter becslésének.

A fentebb leírt eljárás végén adódó paraméterrel aztán megkaphatjuk a populációra a megállók közötti utasszám becslését, és ha minden egyes populációra összeadjuk, akkor megkapjuk a járatra a megállók közötti utasszámot.

6.2. Átmeneti valószínűségek számításának más lehetőségei

A fentebb leírt eljárással számolt paraméter azt is megadja, hogy a j -edik megállóban felszállt utasok közül milyen arányban szállnak le az ℓ -edik megállóban, más szóval, hogy milyen valószínűséggel utazik egy utas a két megálló között. Ezt a valószínűséget tekinthetjük a megállók közötti átmeneti valószínűségnek. Számíthatjuk minden egyes (j, ℓ) megállópárra külön-külön. Vagyis a becslési eljárás leírására felhasznált példán egy sztochasztikus folyamat átmeneti valószínűségének számítására is sor került. (Bár nem ez volt a cél!) A leírt számítás csak egyszerűsítve veszi figyelembe, hogy a felszálló megállók között kölcsönhatás van.

Megállók közötti átmenet valószínűségének becslésére való módszerekre az Interneten konkrét, közvetlen példa, irodalom nem volt található. Közvetetten, Markov-lánckokra vonatkozóan az átmeneti valószínűség számítására azonban voltak módszerek. Amennyiben fel lehet állítani valamilyen Markov-lánc modellét a fel- és leszálló utasok megállónkénti alakulására, vagy a populáció egy utasának leszállására (pl. egy lehetőség a számegyenesen egy speciális vándorlás), akkor lehet használni ezeket a becslési eljárásokat.

Véges állapotú és homogén Markov-láncre az [1], [11] ad meg maximum likelihood becslést az átmeneti mátrix elemeinek becslésére. Esetünkben azonban megállóktól függően változik az átmeneti valószínűség, ezért nemhomogén Markov-lánccal kellene számolni az átmeneti mátrix számításánál. A nemhomogén Markov-láncre az [5] irodalom ad számítási módszert. Amennyiben figyelembe vesszük a felszálló megállók közötti kölcsönhatást, vagyis egyszerre vesszük figyelembe az összes megálló felszálló utasait, akkor többváltozós Markov-lánccal kellene modellezni a folyamatot [3], [6], [13], [19].

Bármelyik esetet nézzük azonban, magát a Markov-láncot, vagy annak egy részét kellene megfigyelni az átmeneti valószínűségi mátrix kiszámításához, vagyis esetünkben a tényleges utasszám alakulását megállónként. De hát éppen ezt szeretnénk tudni! Mivel a folyamatból csak mintával rendelkezünk, így nem alkalmazhatók a Markov-láncre és más idősorokra vonatkozó módszerek.

Ha az átmeneti valószínűség

$$P_{j\ell} = \text{Val}\{(j, \ell) \text{ között egy utas leszállásának valószínűsége}\},$$

akkor az említett szakirodalmakból a számítások általánosan a megfigyelt állapot-

átmenetek arányával fejezik ki ezt a valószínűséget, vagyis

$$P_{j\ell} = \{\text{a } j\text{-ből } \ell\text{-nél leszálló átmenetek száma} / \text{az összes } j\text{-ből származó átmenetek száma}\}.$$

Amennyiben a mintában kapott átmenetek számával, amely most a megállók közötti mintabeli gyakoriságokat jelenti, helyettesítjük a fenti összefüggésben az állapotátmenetek számát, nem várhatunk jó eredményt. Kicsi, vagy nincs is minden (j, ℓ) párra minta. Ezért a $P_{j\ell}$ mátrix erősen pontatlan és ritka lesz. Nem tudjuk, hogy ha nincs mintaelem valamelyik leszálló megállóra, akkor az mit is jelent valójában: kevesen szállnak-e le itt, vagy egyáltalán nincs leszálló utas. Továbbá a minta nem is fejezi ki, hogy a megállók konkurálnak egymással. A mintanagyság növelhető több járat mintájának összevonásával. Azonban a nagyobb mintában levő arányok nem alkalmazhatóak külön-külön az egyes járatokra, mert a járatok forgalma különböző, és a megállók közötti forgalom is máshogy alakul járatonként. Összevont mintának a megállók közötti forgalom számítására történő alkalmazásával a [22] foglalkozik.

7. Egyszerűsítések a gyakorlati esetekben

A célforgalmi minta adatai és a keresztmetszeti adatok minden számításhoz szükséges adatot tartalmaznak, ezért a számítás elvégzése lehetséges. Gyakorlati esetekben azonban számolnunk kell azzal, hogy az adatok nem mindig pontosak, valamint egy járatnak sok megállója is lehet. A gyakorlati esetekben minden egyes mintavételt nem lehet mindig ahhoz a szakaszhoz hozzárendelni, ahol is a mintavétel ténylegesen történt. Ez részben adatfelvételi pontatlanságból eredhet, részben, ha több megálló közel van egymáshoz, akkor a szakaszok nem különíthetők el jól időben a mintavételeknél. Figyelembe kell venni a számolási igényt is, amely a megállók számával erőteljesen növekszik. Továbbá az adatoknál is sok esetben csak település van megjelölve, és nem a konkrét megálló. A gyakorlati esetre a modell egyszerűsítése:

- a) Megállókat összevonhatunk települési szinten, és csak akkor kezelünk külön egy településen belül egy megállót, ha a megállóba érkező, illetve kiinduló utasszám jelentősebben különbözik a megálló esetében a többitől, vagy nagy az utascsere. Azokat a megállókat, ahol kevés az utasmozgás, más megállókkal össze lehet vonni.
- b) Az egyes mintaelemeket a szakaszokhoz a mintavétel ideje alapján rendeljük hozzá, de megengedünk egy rövid időeltérést, és az ebbe az időtoleranciába eső szakaszok valamelyikéhez rendeljük a mintát gyakorlati megfontolások alapján. (Például ha több az utas, akkor nagyobb lehet a vett minta nagysága.)

8. A módszer alkalmazása egy utasfelmérésre

Az ismertetett módszer valós felmérési adatokon lett kipróbálva. A módszerrel az **utazási reláció** paraméterei lettek megbecsülve, továbbá egy másik módon történt becslés eredményével össze lettek hasonlítva. Szűrőpróbaszerűen – egy-két esetben – megnéztük, hogy az adatokban történő változás mennyire változtatja meg az eredményt.

8.1. Az utasfelmérési példa

Az **EMAH** nevű, EU-finanszírozású, ökomobilitást vizsgáló projekt keretében az osztrák-magyar határon átmenő vasúti utazások lettek felmérve 2013-ban néhány vasúti vonalon tavasszal, valamint nyáron a hét ugyanazon három napján [20]. Minden alkalommal célforgalmi kikérdezés, valamint keresztmetszeti számlálás történt több vonatonál. A célforgalmi kikérdezés tartalmazta – más, utazásra vonatkozó kérdések mellett – az induló településre/vasútállomásra, esetleges átszálló vasútállomásra és végül céltelepülésre/vasútállomásra irányuló kérdéseket is. Fel lett véve a vonalszám, a vonat száma (amely vonaton a kikérdezés történt) és a kikérdezés ideje. Megjegyezzük, hogy most a *település* vagy *vasútállomás* fogalmak az eddig használt *megálló* fogalom értelmében vannak használva, kifejezve, hogy a vasúti megállóhelyről van szó, és hogy a vasúti megállóhely lényegében a települést is meghatározza.

A vonatok a vasúti vonalak egy adott szakaszán lettek felmérve. A felmért szakaszon mintavételre bármely két megálló között sor kerülhetett. A felmérés célja miatt általában és többségében a határt átlépő utasok lettek megkérdezve. A keresztmetszeti felmérés is a vasúti vonal vizsgált szakaszának állomásaira történt. A felmért szakasz minden állomásán meg lettek számlálva a fel- és leszálló utasok, természetesen a vonatokon utazó összes utas, tehát nem csak a határt átlépők.

A számításokat csak egy vasúti vonalra (524-esre) végeztük el, a Bécs irányába menő, illetve a Bécs irányából jövő forgalomra. A vizsgált vasúti vonal megállói sorrendben Bécs irányába:

Deutschkreutz
Sopron
Loipersbach-Schattendorf
Marz-Rohrbach
Mattersburg
Mattersburg Nord
Wiesen-Sigleß
Bad Sauerbrunn
Neudörf
Katzelsdorf

Wiener Neustadt Hbf.

Wien

A vizsgált szakasz ezen a vonalon Deutschkreutztól Sopronon keresztül Wiener Neustadtig tartott, a Wiener Neustadt – Wien szakaszon célforgalmi kikérdezés és keresztmetszeti számlálás nem volt. Ennek ellenére a Wiener Neustadt – Bécs közötti utolsó szakasz forgalmára a keresztmetszeti felmérésből adódik érték, mivel közbenso megálló nincs. A módszer alkalmazásánál a fenti állomások csak Soprontól kezdődően érdekesek.

8.2. Az adatok megfeleltetése a módszer alkalmazásához

A feltett kérdés: a vasúti vonal ausztriai állomásain hány utas száll le, akik Magyarországról jönnek, illetve az ausztriai állomásairól hány utas indul, akik Magyarországra utaznak.

A módszerhez a felmért adatokat meg kellett feleltetni. Mivel a kikérdezés gyakorlatilag csak a határátlépő utasokra szorítkozott, úgy tekinthetjük, hogy a mintavétel a határátlépő utasokból történt, vagyis a vonaton levő határátlépő utasok szolgálnak alapsokaságként a mintavételhez. Viszont a forgalomszámlálási adat az összes utasra vonatkozik, tehát minden két szomszédos megálló közötti szakaszra meg kell határozni a vonaton levő határátlépő utasok számát. Ezt az összes utasra vonatkozó keresztmetszeti adatokból valamiféle ésszerű feltételezéssel kaphatjuk meg. A feltett kérdés megfelel annak, vajon a határon (Sopronban) felszálló utasok közül milyen arányban szállnak le, illetve a Sopronban leszálló utasok milyen arányban szálltak fel az egyes osztrák állomásokon. Tehát Sopronban felszállt, illetve leszállt utasok adják a populációt, amelyre a paramétereket (arányokat) számítjuk. Vagyis esetünkben most a két halmaz (a mintavételül szolgáló alapsokaság, illetve a populáció) egybeesik.

A módszer alkalmazásának kipróbálására csak egy egyszerűsített eset lett vizsgálva. Azért, hogy minél kevesebb szakaszra kelljen a határátlépő utasok számát megbecsülni, amely becslés a számítás eredményét befolyásolja, és ezáltal az eredmények összevetését más módon számolt értékekkel megnehezíti, a lehető legkevesebb megálló használata a célszerű. Ehhez, amennyire lehetséges, megállókát kellene összevonni, aminek következményeként pedig az összevont keresztmetszeti utasszámok eredményre vonatkozó hatásával kellene számolni.

Szerencsére az expresszvonatok lehetőséget adtak az említett eljárások elkerülésére, ugyanis e vonatoknál csak 4 megállót kell figyelembe venni. Ezek az alábbiak:

Sopron

Mattersburg

Wiener Neustadt Hbf.

Wien

Ez pedig azt jelenti, hogy csak egy szakaszra kell a határátlépő utasok számát megbecsülni, és pedig Mattersburg – Wiener Neustadt szakaszra, mert a Wiener Neustadt – Wien szakaszon a mintavétel hiányában az értéke közömbös, a Sopron – Mattersburg szakaszon pedig adott. Egyúttal pedig, mivel a megállók időben eléggé távol esnek egymástól, az adatfelvétel ideje alapján, ha csak az nem teljesen hamis, vagy nem hiányzik, az egyes mintákat elég bizonyossággal tudjuk a megfelelő szakaszhoz hozzárendelni, ezáltal a mintavétel idejének kisebb pontatlansága nem okoz problémát a hozzárendelésnél.

Az expresszvonatok felmérési és mintavételi adatait összefoglalóan az 1. táblázat mutatja meg. Láthatóan az egy vonatra átlagosan jutó mintaelem-szám nem nagy. A mintavételi arány széles intervallumban változik, így egyes vonatoknál csak 1-2 mintával számolhatunk.

vonatok		utasszám		mintanagyság*		mintavételi arány*		
		összes	határon átlépő	összes	vonat átlag	átlagos	min.	max.
nyár	12	2901	710	109	9,08	0,15	0,04	0,20
tavaszi	11	1856	797	71	6,45	0,09	0,01	0,60
összes	23	4757	1507	180	7,83	0,12	0,01	0,60

*A határátlépő utasokra

1. táblázat. A vizsgált expresszvonatok utasfelmérésének összesítése

8.3. Az összehasonlító módszer ismertetése

A tárgyalt módszerrel kapott eredményeket összehasonlítottuk másik módon becsült értékekkel. Az összehasonlító számításnál az alapsokaságot szintén a vonaton utazó összes határátlépő utas adja, és a becslést is ugyan erre a halmazra tesszük, vagyis a soproni fel-, vagy leszálló utasokra. Vonatonként ismert a határátlépő utasok száma és a vett minta nagysága, és így a mintavételi arány is.

Az utasszám összehasonlító értékét alapvetően az osztrák állomásokon leszálló soproni utasoknak, illetve a Sopronban leszálló osztrák állomásokról jövő utasoknak a mintában levő arányával számoljuk ki. Ez a 4.1. pontban leírt szokásos arányos becslési eljárás, vonatonként vettük az állomások utasszámainak mintában levő arányait. Figyelmen kívül hagytuk, hogy a mintában levő egyes mintaelemek mekkora sokaságból kerültek ki.

A fenti eljárást önkényes, de ésszerű megfontolással „finomítottuk” is. Aminek a lényege, hogy gyengébb minta esetén az utasszámot nem egyetlen vonat alapján számítjuk, hanem az összes hasonlóan közlekedő vonat figyelembe vételével. Feltételezzük, hogy egy gyengébb minta esetén az egy megállóban fel-, vagy leszállt utasok számát pontosabban határozhatjuk meg, ha az összes hasonlóan közlekedő vonat mintáját is figyelembe vesszük. Megosztottuk a vonat utasszá-

mát: egy részét a vonatra kapott mintában levő arány alapján, a másik részét az összes hasonlóan közlekedő vonat figyelembe vételével számolt arány alapján osztottuk el a megállókra.

A 2. táblázat mutatja azokat a vonat mintavételére vonatkozó számokat, amelyek alapján az utasszámot megosztjuk a kétféle számításhoz.

szétosztás típusa	mintavételi arány	mintanagyság	becslés vonat alapján [%]	becslés összes vonat alapján [%]
a	$\geq 60\%$	közömbös	100	0
b	$\geq 30\%$	és >12	100	0
b	$\geq 25\%$	és >14	100	0
c	közömbös	$< 0,06 * N$	40	60
d	egyéb esete		60	40

Megjegyzés: N a vonat összes utasa

2. táblázat. Vonat forgalmának megosztása a kétféle becslési mód között a vonat mintavétele alapján

Ha a mintavételi arány nagyobb 60%-nál, akkor csak a mintabeli arányt alkalmazzuk az utasszámmra. („a” eset). Ha a mintavételi arány $> 30\%$ -nál, és a mintanagyság 12-nél nagyobb („b” eset), akkor is a mintabeli arányt alkalmazzuk. Ha nagyon gyenge a mintavétel („c” eset), akkor az adott vonat forgalmának 60%-át osztjuk el az összes vonat forgalmának figyelembe vételével, és 40%-át a mintában levő arány szerint számoljuk. Például ha 1 db mintaelem van, vagyis 1 fő utas, akkor a vonat forgalmának 0,4-szerese lesz a mintában levő megálló forgalma, és 0,6-szerese oszlik el az összes megálló között, tehát három megállóra ez a forgalom az összes hasonlóan közlekedő vonat figyelembevételével számolt arányok alapján oszlik el. (Mert három megálló van Sopronon kívül!)

A határszámok önkényesek. Egyrészt úgy lettek megválasztva, hogy egy vonatra a „jó” minta a becsléshez elfogadható pontosságot adjon. Az a) típus esetén ez nyilvánvaló. A b) típusnál, megközelítve valamennyire a valóságot, visszatevés nélküli mintavétellel számolva a legkedvezőtlenebb mintavétel esetén mindkét esetben $\sim 0,15$ lesz a standard hiba felső értéke a paraméterre. Továbbá figyelembe véve az összes vonatot (tehát nem csak az itt vizsgált expresszvonatokat) utasszám és mintavétel tekintetében, az is szempont volt, hogy a választott értékek igazodjanak a vonatok általános mintavételi jellemzőihez: se túl engedékenyek, se túl szigorúak ne legyenek. Szempont volt az is, hogy egy megállóra átlagosan mennyi utas jut. (Ha 1-nél kevesebb, akkor azt gyengébb mintának tekintettük.) Az átlagos megállószaám 10, vagyis 1-2 mintaelem várható átlagosan megállónként b) esetben, ha a határszámnál kicsit nagyobb a mintaelemszám. Az expresszvonatoknál a helyzet valamivel kedvezőbb, ezért a b eset jó mintavételnek vehető

mind a mintavételi arány, mind az átlagosan egy megállóra eső mintaelemszám szempontjából.

Az összehasonlító számításhoz használt mintabeli arány valószínűségeloszlását adott vonaton nem ismerjük. Így a becslés pontosságát nem tudjuk pontosan megmondani, különösen, ha az említett korrigálást is használjuk. Azt azonban kijelenthetjük általánosan, hogy a) és b) esetben a minta jobb, mint a c) és d) esetekben, és ezért pontosabb eredményt várunk az előbbieknél. Ez nem jelenti azt, hogy d) esetben gyakorlati szempontból a kapott becslés nem lehet elfogadható pontosságú.

8.4. Eredmények

8.4.1. A becsült utasszámok

Meghatároztuk a három osztrák állomásra a paramétereik értékét, vagyis a soproni felszálló, illetve leszálló utasokból való részesedésüket - mind a tavaszi, mind a nyári két vizsgált napra. Az ezekkel kapott utasszámok tört értékek, amelyek végül egész értékre lettek kerekítve. A 3. táblázat mutatja az eredményeket. Feltüntettük az egyes szakaszokon a mintavételre vonatkozó jellemző értékeket is, a populáció (együttal az alapsokaság) nagyságát.

Az is látszik a táblázatból, hogy a legtöbb esetben a mintavételek száma kicsi. 12 mintaelemből álló, vagy ennél nagyobb minta 7 vonaton van, 5 elemű, vagy kisebb mintanagyság szintén 7 vonaton van. Ez utóbbi esetben az egy állomásra jutó átlagos mintaelemszám 2 alatt van. Ezért az összehasonlító számításnál - de a leírt módszernél is - nagyobb pontatlanságra számíthatunk.

Feltüntettük a leírt módszerrel számolt paraméter várható értékét és a standard hibáját (mint előzőekben említettük, ezek számolhatók). Egyes esetekben a standard hibát a paraméter lehetséges legnagyobb és lehetséges legkisebb érték különbségének (terjedelemnek) harmadával tettük egyenlővé. Amennyiben egy megállóra nincs vagy fel-, vagy leszálló utas, akkor - értelemszerűen - minden érték 0 a táblázatban. Ha Sopron és a vizsgált megálló között nem volt mintavétel, akkor a *szükséges legkisebb*, valamint a *lehetséges maximális* értékek átlaga lett a becsült érték, és ugyanez az érték lesz a várható érték is, a standard hiba pedig a két érték különbségének harmada lesz. A standard hiba ténylegesen 0, ha csak egyféle leszálló esemény következhet be egy megállóban. Vagyis, bár volt mintavétel Sopron és a vizsgált megálló között, csak 0 leszálló utas volt a mintában. A hibának akkor is 0 értéket adtunk, amikor számított értéke kisebb lett 0,001-nél. Amennyiben a paraméter értéke csak nagyon kis intervallumon belüli értéket vehet fel (0,05-on belül), akkor a várható értéket a számított paraméter értékkel tettük egyenlővé, és a standard hiba értékének a terjedelem 3-ad részét adtuk.

vonat	felhívás napja	vonalszakasz	le/felészálló megálló	Sopronban szakaszon fel/leszállók utasszám* ^a	minutavé- tel száma	para- méter	várható standard hiba	becsült fel/leszálló utasszám	minutavé- tel típusa	utasszám össze- hasonlító módszerrel
1810	tavaszkedd	Sopron - Mattersburg	Mattersburg	36	36	1	0,00	0,00	d	1
1810	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	36	35	2	0,23	0,20	d	28
1810	tavaszkedd	Wiener Neustadt - Wien	Wien	36	32	0	0,77	0,77	d	7
1812	tavaszkedd	Sopron - Mattersburg	Mattersburg	104	104	6	0,03	0,03	d	2
1812	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	104	93	2	0,27	0,23	d	42
1812	tavaszkedd	Wiener Neustadt - Wien	Wien	104	77	0	0,68	0,66	d	61
1814	tavaszkedd	Sopron - Mattersburg	Mattersburg	20	20	10	0,00	0,00	a	0
1814	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	20	10	2	0,42	0,39	a	8
1814	tavaszkedd	Wiener Neustadt - Wien	Wien	20	3	0	0,58	0,52	a	12
1810	tavaszkedd	Sopron - Mattersburg	Mattersburg	32	32	4	0,00	0,00	d	1
1810	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	32	28	1	0,24	0,21	d	6
1810	tavaszkedd	Wiener Neustadt - Wien	Wien	32	20	0	0,76	0,72	d	26
1812	tavaszkedd	Sopron - Mattersburg	Mattersburg	91	91	5	0,00	0,00	d	9
1812	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	91	86	3	0,29	0,26	d	57
1812	tavaszkedd	Wiener Neustadt - Wien	Wien	91	73	0	0,71	0,71	d	25
1814	tavaszkedd	Sopron - Mattersburg	Mattersburg	134	134	6	0,01	0,01	d	10
1814	tavaszkedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	134	128	6	0,41	0,39	d	44
1814	tavaszkedd	Wiener Neustadt - Wien	Wien	134	44	0	0,55	0,47	d	81
7755	tavaszkedd	Wien - Wiener Neustadt	Wien	91	49	0	0,58	0,56	c	71
7755	tavaszkedd	Wiener Neustadt - Mattersburg	Wiener Neustadt	91	90	1	0,41	0,43	c	16
7755	tavaszkedd	Mattersburg - Sopron	Mattersburg	91	90	0	0,006	0,004	c	4
7761	tavaszkedd	Wien - Wiener Neustadt	Wien	70	49	0	0,52	0,62	d	25
7761	tavaszkedd	Wiener Neustadt - Mattersburg	Wiener Neustadt	70	70	5	0,48	0,38	d	43
7761	tavaszkedd	Mattersburg - Sopron	Mattersburg	70	65	1	0,00	0,00	d	2

* A minutavételi szolgáló utasszám

3. táblázat. Vonatonként egyes megállók becsült fel- illetve leszálló utasszáma és az összehasonlító módszerrel kapott értékek 1/3

vonat	felmérés napja	vonalszakasz	le/felszálló megálló	Sopronban fel/leszállók utasszám*	szakaszon mintavétel száma	paraméter méter	várható érték	standard hiba	becsült fel/leszálló utasszám	mintavétel típusa	utasszám össze hasonlító módszerrel	
* A mintavételről szolgáló utasszám												
7749	tavaszi, péntek	Wien - Wiener Neustadt	Wien	95	69	0	0,69	0,72	0,07	66	c	68
7749	tavaszi, péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	95	91	4	0,11	0,17	0,06	10	c	18
7749	tavaszi, péntek	Mattersburg - Sopron	Mattersburg	95	91	2	0,04	0,04	0,01	4	c	19
7755	tavaszi, péntek	Wien - Wiener Neustadt	Wien	56	42	0	0,83	0,81	0,00	46	d	48
7755	tavaszi, péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	56	50	5	0,07	0,09	0,00	4	d	6
7755	tavaszi, péntek	Mattersburg - Sopron	Mattersburg	56	51	2	0,07	0,08	0,00	4	d	2
7761	tavaszi, péntek	Wien - Wiener Neustadt	Wien	57	42	0	0,69	0,75	0,05	39	c	29
7761	tavaszi, péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	57	53	3	0,24	0,19	0,00	14	c	25
7761	tavaszi, péntek	Mattersburg - Sopron	Mattersburg	57	54	0	0,04	0,04	0,03	2	c	3
1810	nyár,kedd	Sopron - Mattersburg	Mattersburg	44	44	6	0,00	0,00	0,00	0	d	0
1810	nyár,kedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	44	38	0	0,21	0,17	0,02	9	d	20
1810	nyár,kedd	Wiener Neustadt - Wien	Wien	44	37	0	0,79	0,78	0,005	35	d	23
1812	nyár,kedd	Sopron - Mattersburg	Mattersburg	75	75	15	0,02	0,02	0,01	2	b	0
1812	nyár,kedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	75	58	0	0,39	0,33	0,06	29	b	30
1812	nyár,kedd	Wiener Neustadt - Wien	Wien	75	44	0	0,60	0,56	0,04	45	b	45
1814	nyár,kedd	Sopron - Mattersburg	Mattersburg	97	97	15	0,03	0,03	0,01	3	d	4
1814	nyár,kedd	Mattersburg - Wiener Neustadt	Wiener Neustadt	97	80	0	0,23	0,18	0,04	22	d	29
1814	nyár,kedd	Wiener Neustadt - Wien	Wien	97	65	0	0,73	0,72	0,01	71	d	64
1810	nyár,péntek	Sopron - Mattersburg	Mattersburg	30	30	3	0,00	0,00	0,00	0	d	0
1810	nyár,péntek	Mattersburg - Wiener Neustadt	Wiener Neustadt	30	27	0	0,08	0,08	0,00	2	d	4
1810	nyár,péntek	Wiener Neustadt - Wien	Wien	30	25	0	0,92	0,91	0,02	28	d	26
1812	nyár,péntek	Sopron - Mattersburg	Mattersburg	77	77	14	0,01	0,01	0,01	1	d	0
1812	nyár,péntek	Mattersburg - Wiener Neustadt	Wiener Neustadt	77	63	0	0,30	0,25	0,05	23	d	24
1812	nyár,péntek	Wiener Neustadt - Wien	Wien	77	48	0	0,70	0,65	0,05	54	d	54

* A mintavételtől szolgáló utasszám

3. táblázat. Vonatonként egyes megállók becsült fel- illetve leszálló utasszáma és az összehasonlító módszerrel kapott értékek 2/3

vonat	felmérés napja	vonalszakasz	le/felszálló megálló	Sopronban fel/leszállók utasszáma*	szakaszon mintavétel száma	paraméter	várható érték	standard hiba	becsült fel/leszálló utasszám	mintavétel típusa	utasszám össze hasonlító módszerrel
1814	nyár.péntek	Sopron - Mattersburg	Mattersburg	100	100	12	0,01	0,003	1	d	1
1814	nyár.péntek	Mattersburg - Wiener Neustadt	Wiener Neustadt	100	88	0	0,19	0,02	19	d	39
1814	nyár.péntek	Wiener Neustadt - Wien	Wien	100	78	0	0,80	0,004	80	d	62
7749	nyár.kedd	Wien - Wiener Neustadt	Wien	41	25	0	0,37	0,14	15	d	19
7749	nyár.kedd	Wiener Neustadt - Mattersburg	Wiener Neustadt	41	37	0	0,12	0,09	5	d	5
7749	nyár.kedd	Mattersburg - Sopron	Mattersburg	41	41	6	0,11	0,01	5	d	18
7755	nyár.kedd	Wien - Wiener Neustadt	Wien	70	50	0	0,72	0,07	50	c	35
7755	nyár.kedd	Wiener Neustadt - Mattersburg	Wiener Neustadt	70	64	0	0,13	0,04	9	c	12
7755	nyár.kedd	Mattersburg - Sopron	Mattersburg	70	70	3	0,09	0,08	6	c	23
7761	nyár.kedd	Wien - Wiener Neustadt	Wien	50	25	0	0,41	0,12	21	d	20
7761	nyár.kedd	Wiener Neustadt - Mattersburg	Wiener Neustadt	50	36	0	0,52	0,08	26	d	24
7761	nyár.kedd	Mattersburg - Sopron	Mattersburg	50	50	10	0,08	0,14	4	d	7
7749	nyár.péntek	Wien - Wiener Neustadt	Wien	27	5	0	0,10	0,06	3	d	6
7749	nyár.péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	27	22	1	0,42	0,51	11	d	7
7749	nyár.péntek	Mattersburg - Sopron	Mattersburg	27	26	5	0,24	0,23	6	d	23
7755	nyár.péntek	Wien - Wiener Neustadt	Wien	50	42	0	0,79	0,85	40	d	17
7755	nyár.péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	50	49	4	0,17	0,01	9	d	24
7755	nyár.péntek	Mattersburg - Sopron	Mattersburg	50	46	9	0,03	0,01	2	d	10
7761	nyár.péntek	Wien - Wiener Neustadt	Wien	36	14	0	0,19	0,09	7	d	14
7761	nyár.péntek	Wiener Neustadt - Mattersburg	Wiener Neustadt	36	34	0	0,58	0,57	21	d	8
7761	nyár.péntek	Mattersburg - Sopron	Mattersburg	36	36	5	0,06	0,06	2	d	14

* A mintavételül szolgáló utasszám

3. táblázat. Vonatonként egyes megállók becsült fel- illetve leszálló utasszáma és az összehasonlító módszerrel kapott értékek 3/3

8.4.2. Az összehasonlítás eredménye

A 3. táblázat az ismertetett másik módszerrel való összehasonlítás eredményét is mutatja. Amikor csak a mintabeli aránnyal számoltuk a megállók forgalmát (a és b típusú eseteknél), csupán 2 eset volt ilyen, akkor jó egyezés van a két becslési eredmény között. Amikor gyengébb volt a minta (d eset), és ez volt a legtöbb, akkor esetenként hol egyezés volt, hol jelentősebb eltérés adódott. Kb. azonos számban, és a mintavételi arány sem látszik perdöntőnek. Most nem bemutatva az eredményeket, el lett végezve a számítás csak a mintabeli arány használata alapján (finomítás nélkül), de az se adott jobb egyezést. Ha nagyon gyenge volt a minta (c eset), akkor az eredmények meglehetősen eltérnek. (A megadott standard hiba 2-szeresénél is, sőt a 3 szorosánál is vannak nagyobb eltérések.)

Nyilvánvaló volt a tapasztalatok alapján, és ez az eredményekből látszik is, hogy általában a Bécsbe, illetve onnan utazók aránya a legnagyobb, olykor eléggé dominálón. Ez még akkor is így van, amikor nincs is olyan minta, amely Bécsben felszálló, vagy leszálló utasra vonatkozik. Ezért tűnik elfogadhatóbbnak például a tavaszi keddi 1810-es vonatnál a Bécsben leszállók számának az általunk használt módszerrel kapott értéke (28), mint az összehasonlító eljárással számolté (7). A bécsi utasszámra egy-két ilyen jelentősebb eltérés még előfordul az eredmények között.

Megjegyzés: Esetenként az összehasonlító módszernél figyelembe vett alaposság kicsit eltér a táblázati értéktől, mert a Deutschkreutz – Sopron közti utasok is benne vannak. Ez azonban nem befolyásolja az eltérés megítélését.

8.4.3. Érzékenységi vizsgálat

Megvizsgáltuk, hogy a számítás eredményét a pontatlan, hibás adatok mennyire befolyásolják. Nem törekedtünk teljes, megalapozott vizsgálatra, csak egy-két esetben néztük meg az eltérést.

A fel- és leszállók számát, és így a szakaszonkénti utasszámot pontosnak lehet tekinteni, hiszen csak egy szakaszra lett utasszám becsülve. A mintánál pedig elfogadhatjuk, hogy hol szállt fel, vagy le az utas. A leginkább problémás adat, – és ennek több oka is van – ha nem ahhoz a szakaszhoz rendeljük a mintavételt, ahol az ténylegesen megtörtént. Ezért megvizsgáltuk, hogy ha a mintavétel egyes mintaelemeknél más szomszédos szakaszon történik, hogyan befolyásolja az eredményt. Olyan esetekben érdemes ilyet nézni, ahol egy szakaszon több mintaelem lett véve.

Ezért megváltoztattuk az egyes szakaszokon vett mintaelemek darabszámát, de meghagyva a megállóra vett összdarabszámot. (A 4.3. pontban leírtak szerint az érzékenységet vizsgálni lehet bootstrap módszerrel is, mi nem ezt az utat választottuk.) A 4. táblázat bemutatja a kapott eltérést.

A táblázat tartalmazza az eredeti számításnál használt szakaszonkénti utasszámot, és a szakaszonkénti mintavételek számát, valamint ugyanezen adatokat a min-

vonat	felmérés napja	vonalszakasz	le/felszálló megálló	populáció nagysága	szakaszon utasszám*	eredeti mintavételi szám	szakaszon módosított utasszám*	módosított mintavétel i szám	eredeti számolt paraméter	módosítás utáni paraméter	utasszám változás
* A mintavételi szolgáló utasszám											
1812	nyár,kedd	Sopron - Mattersburg Mattersburg - Wiener Neustadt Wiener Neustadt - Wien	Mattersburg Wiener Neustadt Wien	75 75 75	75 58 44	15 0 0	75 63 44	10 5 0	0,02 0,39 0,60	0,018 0,386 0,596	0 0 0
1814	nyár,kedd	Sopron - Mattersburg Mattersburg - Wiener Neustadt Wiener Neustadt - Wien	Mattersburg Wiener Neustadt Wien	97 97 97	97 80 65	15 0 0	97 87 65	8 7 0	0,03 0,23 0,73	0,028 0,224 0,727	0 -1 0
7761	nyár,kedd	Wien - Wiener Neustadt Wiener Neustadt - Mattersburg Mattersburg - Sopron	Wien Wiener Neustadt Mattersburg	50 50 50	25 36 50	0 0 10	25 36 47	0 3 7	0,41 0,52 0,08	0,375 0,499 0,097	-2 -1 1
1814	nyár,péntek	Sopron - Mattersburg Mattersburg - Wiener Neustadt Wiener Neustadt - Wien	Mattersburg Wiener Neustadt Wien	100 100 100	100 88 78	12 0 0	100 92 78	8 4 0	0,01 0,19 0,80	0,005 0,192 0,798	-1 0 0
7755	nyár,péntek	Wien - Wiener Neustadt Wiener Neustadt - Mattersburg Mattersburg - Sopron	Wien Wiener Neustadt Mattersburg	50 50 50	42 49 46	0 4 9	42 49 44	0 6 7	0,79 0,17 0,03	0,793 0,171 0,031	0 0 0

4. táblázat. A szakaszonkénti mintavételi számra való érzékenység vizsgálata egyes vonatoknál

tavételek egy részének más szakaszokhoz történt hozzárendelése esetén. A paramétereket mindkét esetre kiszámoltuk. Látható, hogy jelentősen nem változik a számolt utasszám, pedig a mintaelemek egy része más szakaszra került át. Ez az eredmény részben várható is, hiszen a súlyozott átlag értéke az eloszlás változásával csak kevésbé változik. Az eloszlás változása nagyobb mértékű lehet, ha a mintavételül szolgáló utasszám a szakaszokon kisebb érték. Erre vizsgálatot nem végeztünk.

9. Összegzés

A cikkben bemutatunk egy két lépésből álló, a szokásos eljárásoktól valamegyest más megközelítésű becslési módszert arra vonatkozóan, hogyan becsülhetjük meg kategória változók populációbeli valószínűségeloszlását megadó paramétereket bizonyos mintavételi adottságok esetén. Az első lépésben a vizsgált kategória mintában levő gyakoriságának eloszlását határozzuk meg a paraméterre felvett valamilyen érték mellett. A konkrét számítási eljárás függ a mintavételi módtól. A második lépésben a paramétert becsüljük meg a gyakoriság eloszlásának számításához használt különböző paraméter értékek súlyozott átlaga alapján, ahol a súlyok a gyakoriság eloszlásából kaphatóak meg. Ezt a módszert leginkább az utasforgalom felméréséből történő becslésnél látjuk alkalmazhatónak az eddig szokásos arányos becslés helyett, főleg olyan esetekben, amikor a minta kicsi. A módszer matematikailag mindenképpen megalapozottabbnak látszik, mint az eddig használt módszer, és akár standard hibát, akár konfidencia intervallumot lehet számolni a módszerhez kapcsolódóan kialakított modell keretein belül, ismerve a vizsgált kategória gyakoriságának mintabeli eloszlását.

A súlyozott átlaggal számolt becslési módra nem találtunk irodalmat, ezért a statisztikai tulajdonságait a későbbiekben vizsgálni tanácsos akár elméletileg, akár generált minták alapján. A cikkben erre nem tértünk ki, nem is történtek ilyen irányú vizsgálatok, számítások.

Az írás célja volt az is, hogy a módszer alkalmazása a számítási eljárás járművön történő felmérés esetére be legyen mutatva, valamint gyakorlati példán az általa kapott eredmények is.

A kapott eredményeket az alkalmazott összehasonlító módszerrel összevetve az látszik, hogy jónak vehető minta esetén a kétféle módon kapott értékek közel esnek egymáshoz. Gyengébbnek tartott mintánál egyes esetekben jelentősebb eltérés is adódik. Mivel jó mintáknál a két eredmény közel esett egymáshoz, ezért más minták esetén is feltételezhetjük, hogy a módszerrel kapott eredmények elfogadhatók, különösen akkor, ha a standard hiba kis értéknek adódik. (A mintában a megállók leszálló utasszámainak rögzítése miatt a becsült értékre ugyan kisebb standard hiba adódik, de ennek ellenére azt iránymutatónak gondoljuk.) Az eredmények alapján erősen állítható, hogy a módszer ugyanolyan, sőt pontosabb eredményt

ad, mint a szokásos eljárás, ráadásul akkor is kapható becslés ez alapján, ha nem volt mintavétel a vizsgált megállók között.

A bemutatott módszer nagyobb számolási igényű, mint egy egyszerű felszorzás valamilyen aránnyal, vagy akár valamilyen heurisztikus eljárás alkalmazása. Ha nem is mindegyik megállóra, de a fontosabb megállóknál alkalmazásra javasolható.

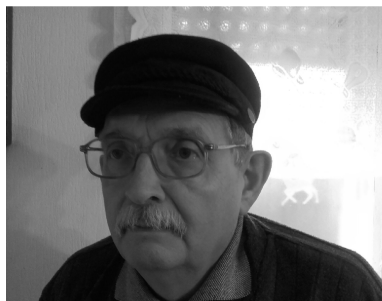
Nem gondoljuk, hogy a modellben a mintavételnek teljesen a valóságban történtek szerint kell lennie. A szűrőpróbaszerű érzékenység vizsgálat alapján úgy tűnik, hogy nem kell pontosan a valóságban történt mintavételt leutánozni, csak egy, a ténylegeshez közeli helyzetet kell leírni, és arra végezni el a számítást.

Hivatkozások

- [1] BÖJTHY BARBARA ADRIENN: *Sztochasztikus mátrixok és Markov-láncok*. Bsc szakdolgozat (2013), www.cs.elte.hu/blobs/diplomamunkak
- [2] YONGYUTH CHAIYAPONG: *Estimation of Proportion of a Finte Population* 2012. **10(1)** Thailand Statistician
- [3] WAI-KI CHING, SHU-QIN ZHANG, AND OTHERS: *On Multi-dimensional Markov Chain Models*, www.hkumath.hku.hk/~imr/IMRPreprintseries/2006/IMR2006-21.pdf, Pacific Journal of Optimization 2007
- [4] WARRENT DENT, RICHARD BALLINTINE: *A Review Of Estimation of Transition probabilities In Markov Chains*. The Australian Journal Of Agriculture Economics Vol. **15** No. **2**, (1971)
- [5] T.R. FLEMING – D.P. HARRINGTON: *Estimation for Discrete Time nonhomogeneous Markov Chains*. Stochastic Processes and their Applications **7**, (1978) 131–139.
- [6] SU-LIEN LU: *Using a Multivariate MarkovChain Model For Estimating Credit Risk: Evidence from Taiwan*, Advances in Managment & Applied Economics Vol. **2** No. **4**, (2012).
- [7] YUTONG LI, HONGXING YANG: *First Order Multivariate Markov Chain Model for* International Refrigeration and Air Conditoning Conference at Purdue (2008), <http://docs.lib.purdue.edu/iracc>.
- [8] STEVEN STERN: *Simulation-Based Estimation and Inference and Random Parameter models*. www.people.stern.nyu.edu/wgreen/Lugano2013 ????
- [9] STEVEN STERN: *Simulation-Based Estimation*. (1997) www.people.virginia.edu
- [10] TAKÁCS SZABOLCS: *Érzékenységvizsgálatok a statisztikai eljárásokban*. Alkalmazott Matematikai Lapok **29** (2012).
- [11] IULIANA TEODORESCU: *Maximum likelihood Estimation For Markov Chains*. www.arxiv.org/pdf/0905.4131.pdf
- [12] *Hypergeometric distribution*. www.math.uah.edu/stat3/urn/introduction.html
- [13] WAI-KI CHING AND OTHERS: *A New Multivariate Markov Chain Model with Applications to Sales Demand Forecasting*. International Conference on Industrial Engineering and System Managment (2007) Beijing, www.hkumath.hku.hk/~imr/.../2007/IMR2007-17.pdf

- [14] *A new Multivariate Markov Chain Model for Adding a New Categorical Data Sequence.*
<http://www.hindawi.com/journals/mpe/2014/502808>
- [15] SAMUEL KARLIN, HOWARD M. TAYLOR: *Sztochasztikus folyamatok.*
- [16] E.L. LEHMANN: *Elements of Large Sample Theory.* Springer
- [17] *Chapter 1: Bootstrap Method.* www.math.ntu.edu.tw/~hchn/LargeSample/...
- [18] SAJTOS LÁSZLÓ – MITEV ARIEL: *SPSS Kutatási és adatelemzési kézikönyv.* Alinea Kiadó
- [19] WAI-KI CHING, MICHAEL K. NG: *Markov Chains Models.* Algorithm and Applications
Springer (2006) Science + Business Media, www.down.cenet.org/upfile/28/2007.....pdf
- [20] *EMAH – Ökomobilitás elősegítése az osztrák-magyar határtérségben;* tanulmány, Közlekedéstudományi Int. (2014). Témavezető: Virág Álmos
- [21] VASS LAJOS: *Másféle megközelítés utazási jellemzők gyakoriságának becslésében a célforgalmi felmérésből.* Közlekedéstudományi Int., Évkönyv 2011–12.
- [22] VASS LAJOS: *A helyi és helyközi tömegközlekedési forgalom keresztszeti felmérésének konverziója valószínűségi módszerrel.* Közlekedéstudományi Szemle 2001. 01.

(Beérkezett: 2016. szeptember 20.)



Vass Lajos 1948-ban született. Az ELTE-n okleveles fizikusként végzett. A Közlekedéstudományi Intézetnek volt főmunkatársa, most már visszavonult.

1992-től az Intézetben a közlekedési (áruszállítási, személyszállítási) adatok feldolgozásával, kiértékelésével foglalkozott, valamint a kiértékeléshez használható matematikai módszerek, elsősorban matematikai-statisztikai módszerek alkalmazásával, kidolgozásával.

Számos tanulmány elkészítésében közreműködő, illetve társszerző volt. A módszerek alkalmazásával kapcsolatban néhány (5 db) cikket jelentetett meg itthoni lapokban.

Az Intézeti munkájának elismeréséül a „KTI-ért” díjat kapta (kétszer) munkahelyétől, valamint a szakminisztériumtól „Közlekedésért” érmet kapott miniszteri kitüntetésként.

1992 előtt az előző munkahelyein, a Csepel Művek fejlesztési, kutatási és tervező intézeteiben dolgozott. Gyártási hiba keresésére és anyagvizsgálati adatok kiértékelésére szintúgy matematikai-statisztikai módszereket használt (többváltozós lineáris regressziót, clusteranalízist, stb.). Számítógépes modellezést is végzett fizikai folyamatokra tervezési munkák segítéséhez. Spektrometriás kiértékelési

program kifejlesztéséért kollégájával fiatal kutatók akadémiai elismeréseként Erdey Ferenc-díjat kapott.

VASS LAJOS

KTI Közlekedéstudományi Intézet

A PROBABILITY METHOD FOR THE ESTIMATION OF TRIP CHARACTERISTICS FROM PASSENGER SURVEYS

LAJOS VASS

The paper introduces an estimation procedure that is different from the mainstream and widely-used methods for calculating distribution and parameters of categorical variables. This procedure is based on sampling distribution of the frequency of analyzed category, calculated in function of the parameter. The estimator of the parameter is weighted average of parameter values used for calculation, in which weighting factors are the probabilities of the frequency observed in the sample, obtained with different values.

Essential point is modeling sampling and how to calculate distribution on the base of the model. As it has been developed to help the analysis of passenger trip characteristics, its application may be useful for estimations on the basis of passenger surveys. It might be useful especially for small samples. The paper describes the method and calculation techniques for a special survey – questionnaires on-board public transport vehicles – and an application is presented here through a real-life example.