

Bayesi gondolkodás a statisztikában*

Hunyadi László

CSc, egyetemi tanár,
a *Statisztikai Szemle*
főszerkesztője

E-mail: Laszlo.Hunyadi@ksh.hu

A szerző a bayesi gondolkodás és kiemelten a bayesi statisztikai következtetésemélet fontosabb pontjait mutatja be tanulmányában. Bayes eredeti formulájából indul ki, majd rámutat arra, hogy ez miként értelmezhető általánosabb keretekben, és elvezet az általános tudományos gondolkodás egyszerű modelljéhez. A következtetésemélet tárgyalásakor kiemeli azokat a pontokat, amelyek éles választóvonalat jelentenek a bayesi és a klasszikus statisztika között, majd sorra veszi a bayesi következtetésemélet fontosabb elemeit. Különös hangsúlyt kap a prior értelmezése és konstruálása, valamint a posteriorra épülő, utólagos elemek, amelyek összeköttetést biztosítanak a klasszikus és a bayesi eredmények között. A cikk röviden érinti a hipotézisvizsgálat bayesi szemléletű megközelítését, és utal az ezen elemzésekből kinőtt, mára már nagyjelentőségűvé vált algoritmusokra is. A nagyszámú alkalmazásból kiemeli a gazdaságelemzés néhány magyar eredményét. Befejezésül összeveti a statisztika két lehetséges eszköztárát, hangsúlyozva, hogy azok egymást erősítő, nem pedig egymással rivalizáló eredményekre is vezethetnek, ugyanakkor a világ jobb megismerése érdekében mindkét szemlélet alapos ismeretére van a statisztikusoknak szüksége. Ez egyben az oktatás, a kutatás ezzel kapcsolatos feladatait is körvonalazza.

TÁRGYSZÓ:
Bayes-féle becslés.

* A szerző ezúton fejezi ki köszönetét a dolgozat két lektorának, *Kehl Dánielnek* és *Várpalotai Viktornak* hasznos tanácsaikért.

Amikor *Thomas Bayes* az 1700-as évek közepén megírta híres művét¹ az inverz valószínűségekről, aligha gondolta, hogy a XX. és XXI. század tudományára kiható alapokat rakott le. Nem gondolhatta már azért sem, mert valójában mindaz, amit ma nevével jegyeznek, csak nyomaiban emlékeztet az eredeti bayesi gondolatokra. Közismert mondással akár azt is mondhatnánk, hogy Bayes majdnem úgy került be az utóbbi fél évszázadnak szinte az egész tudományos gondolkodását megrázó fejleményeibe, mint Pilátus a Credoba. Ez persze legfeljebb az elnevezést, és Bayes érdemeit érintő megjegyzés lehet, de a lényegét – nevezetesen azt, hogy a XX. század elejéig kialakult, statisztikán és valószínűség-számításon alapuló metodológia mellé felnőtt egy új, azt lényegileg tagadó alternatív módszertan – nem befolyásolja. Mivel ez az irányzat összefoglaló néven Bayes-módszertan, vagy a mi szakterületünk szűkebb értelmezésében bayesi statisztika néven terjedt el, természetesen így használjuk a későbbiekben. A továbbiakban megpróbáljuk bemutatni azokat a jellegzetességeket, gondolati alapjait, az azokkal kapcsolatos vitákat, pró és kontra érveket, amelyek végigkövették és követik ma is ennek az egyre inkább terjedő szemléletnek a történetét és jelenét.

A tanulmány további szerkezete a következő: a bevezetés utáni első fejezetben a Bayes-tétel eredeti formáját, néhány alkalmazását és a mögöttes gondolkodást vázoljuk. A második részben azt mutatjuk meg, hogy a követők az új szemléletű módszertan kialakításakor milyen további értelmezéseket csatoltak az eredeti bayesi gondolatokhoz, és hogyan vált a bayesi elmélet mintegy általános tudományos módszertanná. A harmadik fejezetben már a statisztikára kívánunk közelíteni: lépésről lépésre bemutatjuk a bayesi következtetéselemélet fontosabb elemeit, valamint a hozzájuk kapcsolódó gondolatokat. Ezt követően a bayesi módszerek gyakorlati megvalósítását és lehetséges alkalmazásait ismertetjük. A cikket a módszertan általános értékeléseként is tekinthető összefoglalóval és néhány, a témában releváns és hozzáférhető szakirodalmi hivatkozással zárjuk.

1. Bayes tétele és interpretációja

Bayes eredeti gondolatait mai formalizálással a legegyszerűbb esetben a következőkkel írhatjuk le: legyen A és B két egymástól nem független esemény, ekkor a feltételes valószínűségek definíciójából könnyen adódik, hogy

¹ A művet valójában Bayes halála után tanítványa és barátja, *Robert Price* hozta nyilvánosságra (*Bayes* [1958]).

$$\Pr(AB) = \Pr(A|B)\Pr(B) \text{ és } \Pr(BA) = \Pr(B|A)\Pr(A),$$

és mivel az együttes bekövetkezés valószínűsége szimmetrikus, a jobb oldalak egyenlővé tételével

$$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A),$$

majd átrendezésével azonnal kapható az, amit a valószínűség-számítás Bayes-tételnek nevez:

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}. \quad /1/$$

Amennyiben B nem egyetlen eseményt, hanem B_1, B_2, \dots, B_K az események teljes rendszerét jelenti, akkor a korábbiakhoz hasonlóan származtatható a teljes valószínűség tétele

$$\sum_{i=1}^K \Pr(A|B_i)\Pr(B_i) = \Pr(A),$$

valamint a Bayes-tétel:

$$\Pr(B_j|A) = \frac{\Pr(A|B_j)\Pr(B_j)}{\sum_{i=1}^K \Pr(A|B_i)\Pr(B_i)}, \quad j = 1, 2, \dots, K.$$

A gyakorlatban ez utóbbi forma a fontosabb, de a lényegét az egyszerűbb /1/ is jól mutatja. Ez pedig nem más, mint a valószínűségek felcserélhetősége. Amennyiben a feltételes valószínűségeknek ok-okozati interpretációt adunk, akkor az okok és okozatok egymással való kapcsolatát, az okozati láncon való kétirányú mozgást írja le. Amennyiben előzmény-következmény módon gondolkodunk, akkor időben történő visszafelé következtetés eszközét is láthatjuk a Bayes-tételben. Mindenképpen a fordított irányú gondolkodás (inverz-valószínűségek számítása) az az elem, ami itt lényeges, és ez volt az a mozzanat, ami a klasszikus alkalmazásokat jellemezte. Egy egyszerű, de jellemző példa erre a következő.

1. példa

Legyen azonos alkatrészekből két beszállító! Korábbi nagyszámú vizsgálat alapján tudjuk, hogy A szállító esetén a selejtarány $P = 0,2$, B szállítónál pedig $P = 0,3$. Nem tudjuk biztosan, hogy egy adott szállítmány milyen valószínűséggel, melyik szállítótól származik.

Előzetes feltételezés (szubjektív hit): 0,6 annak valószínűsége, hogy a szállító az A , 0,4 a valószínűsége annak, hogy a szállító a B volt. Ez azon a tapasztalaton alapulhat, hogy a korábbiakban az A és a B szállító ilyen arányban szállított. A sejtés pontosítása mintavétellel: egyetlen kételemű (FAE) mintában egy hibátlan és egy selejtes alkatrészt találtunk. Vizsgáljuk meg, hogyan módosul az előzetes feltételezés (az egyes beszállítóktól érkezés valószínűsége) a minta ismeretében!

A megoldást most csak a becslésre írjuk fel (a hipotézisvizsgálat számításait mellőzzük).

$\theta = P$ paraméter diszkrét, csak két értéket vehet fel, *a priori* valószínűség-eloszlása:

$$\Pr(P = 0,2) = 0,6,$$

$$\Pr(P = 0,3) = 0,4.$$

A selejtes darabok száma (k) binomiális eloszlást követ, ezért a fordított irányú valószínűség (likelihood függvény) azonnal felírható.

$$\Pr(k = 1|P = 0,2) = 2P(1 - P)|(P = 0,2) = 0,32,$$

$$\Pr(k = 1|P = 0,3) = 2P(1 - P)|(P = 0,3) = 0,42.$$

A keresett feltételes valószínűség szintén „kételemű” eloszlás lesz:

$$\Pr(P = 0,2|k = 1) = \frac{0,192}{0,36} = 0,53,$$

$$\Pr(P = 0,3|k = 1) = \frac{0,168}{0,36} = 0,47,$$

ahol a törtek nevezőit a teljes valószínűség korábban idézett tétele alapján határoztuk meg.

A valószínűségek összege természetesen 1-et ad. Ezek azt mutatják, hogy a kezdeti hitünk a minta hatására némileg megváltozott (például az A -tól 0,6-ról 0,53-ra).

A példából jól látható a fordított irányú gondolkodás, ami egyébként nem teljesen idegen a klasszikus statisztikától sem. (Elegendő hivatkozni a maximum likelihood módszerre, amely *olyan paramétereket keres, amelyeket, ha a sokasági eloszlásra érvényesnek tekintünk, biztosítják, hogy aktuális mintánk az összes más paraméterrel rendelkező sokaságból a leginkább hihető módon adódik.*) Ez a fordított irányú logi-

ka ugyanakkor a köznapi gondolkodásban nem természetes. Egy friss tanulmányban (Kadane [2009]) a szerző a korai detektívrodalomból vesz példákat ennek interpretálására: a detektív munkája is hasonló fordított gondolkodást igényel, hiszen fordítva járja be az oksági láncot. A detektív szembesül az eredménnyel (okozat), és azt próbálja megtalálni, hogy milyen okok vezethettek a tapasztalt okozathoz. A cikk idézi ezzel kapcsolatban *Sir Conan Doyle-t*: „A legtöbb ember képes arra, hogy ha az események egy sorát leírjuk, az eredményt megmondják. Össze tudják az agyukban rakni a következményeket, és előre tudják jelezni a várható kimenetet. Kevesen vannak ugyanakkor olyanok, akik, ha elmondunk nekik egy eredményt, fel tudják építeni azt a láncot, látják azokat a lépéseket, amelyek ehhez az eredményhez vezettek.” Végül is éppen ez az, amit a Bayes-tétel formalizál.

Ez az oknyomozó gondolkodásmód kétségtelen igen lényeges eleme a modern tudományos módszertannak, de csak egy eleme. A következőkben azt próbáljuk megmutatni, hogy milyen további elemekkel bővült ez az elmélet napjainkig.

2. A Bayesi gondolkodás, mint általános módszertani keret

Már a korai matematikus gondolkodók – elsősorban *Laplace* – felismerték, hogy a Bayes által megfogalmazott összefüggésekben több van, mint amit közvetlenül leolvashatunk belőle. Lassan, a XX. század közepére kialakult az az elgondolás, amely szerint a Bayes-tétel egyfajta általános értelmezésben a tudományos gondolkodás egyszerűsített modellje lehet, olyan keret, amibe konkrét tartalom függvényében egyszerűsített formában bár, de az egész tudományos módszertan belefér. Ehhez az /1/-ben felírt összefüggés A és B eseményeit egy kicsit át kell fogalmazni.

Legyen most H egy tetszőleges hipotézis, E pedig valamiféle hozzátartozó empirikus bizonyíték. Legyenek továbbá $\Pr(H)$ és $\Pr(E)$ valószínűségek, melyek közül $\Pr(H)$ a H hipotézis bekövetkezésének tapasztalás előtti (a priori) valószínűsége, $\Pr(E)$ pedig az empiria bekövetkezésének valószínűsége (minden szóba jöhető H fennállása esetén). A $\Pr(E|H)$ feltételes valószínűség azt mutatja meg, hogy a prior hipotézis fennállása esetén milyen valószínűséggel kaphatjuk meg az E empirikus bizonyítékot. Míg a fordított irányú feltételes valószínűség, az utólagos (a posteriori) valószínűsége annak, hogy a bizonyíték megtörténte, beszerzése után a H hipotézist fenntartjuk.

Bayes-tétele ekkor

$$\Pr(H|E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)} \quad /2/$$

alakú, ahol a $\Pr(H)$ -t prior valószínűségnek (vagy röviden priornak), a $\Pr(H|E)$ -t posterior valószínűségnek vagy röviden posteriornak nevezzük. Megjegyezzük, hogy több lehetséges H_i hipotézis esetén a nevező a korábban bemutatott teljes valószínűség tétele értelmében a következőképpen írható fel:

$$\Pr(E) = \sum_i \Pr(E|H_i)\Pr(H_i).$$

A /2/ formula központi jelentőségű a bayesi gondolkodás megértésében. Azt mondja ugyanis, hogy kiindulva valamiféle tudásból, tapasztalatokat gyűjtünk be, azokat értékeljük, konfrontáljuk az induló feltételezésünkkel, és ezáltal a tudás egy magasabb szintjére jutunk el. Ebben az értelemben a /2/ valóban a tudományos megismerés nagyon általános és egyszerű modellje. Ugyanakkor van ennek az összefüggésnek egy másik olvasata is, ami nem kevésbé lényeges. Ez pedig az, hogy döntést kell hoznunk bizonytalan környezetben: azt kell eldöntenünk, hogy a H hipotézis, illetőleg annak bekövetkezésére vonatkozó várakozás helyes-e? A kapott (ugyancsak valószínűségi állítás formájában megfogalmazott) empirikus bizonyítékok alapján felülvizsgáljuk eredeti állításunkat, és döntésünket az induló elgondolások (döntési helyzet), a tapasztalatok, valamint ezek valószínűségei alapján hozzuk meg. A sztochasztikus környezetben hozandó döntések általános sémája ugyancsak jól leírható a /2/ összefüggéssel.

Ahhoz azonban, hogy ezeket az összefüggéseket jobban megértsük, legalább két kérdést kell alaposabban megvizsgálni: az egyik a prior tartalma és jelentése, a másik a valószínűség fogalmának a használata. A prior léte, tartalma, értelme az egész bayesi gondolkodás egyik sarokpontja. A klasszikus módszertan (beleértve természetesen a statisztikát is) következtetéseit tapasztalatokból (megfigyelések, mérések stb.) vonja le, hangsúlyozva azok objektív (az elemzést végző személytől független) jellegét. Ezzel szemben a bayesi gondolkodás modellje általánosabb, hiszen azt mondja ki, hogy létezhetnek (sőt szinte mindig léteznek) mintán kívüli információk, amelyek egy része lehet szubjektív is (hiszen az elemzőnek lehetnek előzetes ismeretei a tárgyról), és amelyek az egész tudományos következtetést vagy döntést befolyásoló lényeges tényezők lehetnek. A priorok fontosságának megvilágítására álljon itt két egyszerű példa.²

2. a) példa

Egy zenész azt állítja, hogy egyetlen partitúraoldal áttanulmányozása után megmondja, hogy a nevezett mű *Haydn* vagy *Mozart* alkotása. Állítását ellenőrzendő tíz véletlenszerűen és egymástól függetlenül kiválasztott oldalt mutattak meg neki, és ő mind a tízről helyesen állapította meg, ki a szerző.

² A példák *Savage-tól* származnak, *Wickmann* [1995] idézi őket.

2. b) példa

A kocsmában részeg társaság azzal szórakozik, hogy egy társukat ráveszik, próbálja meg eltalálni: a feldobott érme fej vagy írás lesz-e. Egymás után tízszer elvégzik a kísérletet, és társuk mind a tíz alkalommal helyesen találta el, hogy mi lesz a dobás eredménye.

A szituációk összevetésekor azonnal megállapíthatjuk, hogy amennyiben csak a tapasztalatokat tekintjük, a két eset azonos következtetésre vezet: ha feltételezzük, hogy a személyek mindkét esetben csak véletlenül találgattak, akkor egy ilyen tízes találati valószínűségi sorozat bekövetkezési valószínűsége $P = (1/2)^{10} = 1/1024 \cong 0,001$, azaz nagyon kicsi. Ha csak ezt tekintjük, azt állíthatnánk, hogy a zenész alighanem valóban érti a szakmáját, a részegnek pedig paranormális képességei vannak.

A helyzet ismeretében azonban aligha meglepő, ha nem hagyatkozunk csupán a tapasztalatokra, hanem egyéb külső ismereteinket is igénybe véve, prior valószínűségeket rendelünk a két esethez: míg a zenésznél nagy valószínűséggel hihető, hogy valóban érti a szakmáját, a részegnél paranormális képességnek (előre megmondja a pénzfeldobás eredményét) alighanem igen kis előzetes valószínűséget adunk.³

A prior valószínűség kérdésére a későbbiekben még visszatérünk, de előbb vizsgáljunk meg egy másik kérdést, a valószínűség fogalmát. Míg a klasszikus tudományos felfogásban a valószínűség objektív fogalom: a relatív gyakoriságok háttértéke, a bayesi keretek ennél jóval szabadabb értelmezést is lehetővé tesznek. Ez utóbbi értelmezésben a valószínűség nem (csak) tömegesen előforduló események bekövetkezését jellemző számérték, hanem egyes, nem ismétlődő események vagy állapotok bekövetkezésének várt gyakoriságát jellemző mérőszám. Ha egy olyan eseménynek a bekövetkezését akarjuk jellemezni mint például, hogy ki nyeri 2012-ben a Bajnokok Ligáját, akkor valamiféle szakmai megfontolás, szubjektív érzés, megítélés, hit alapján mondhatjuk azt, hogy a Barcelonának 50 százalék esélye van erre, de jól látható, hogy ennek az 50 százaléknak semmi köze nincs a hagyományosan értelmezett (objektív) valószínűséghez. Az ilyen esetekben azt mondjuk, hogy szubjektív valószínűséget határozzunk meg. Ez a valószínűség nem megalapozott (vagy másképp megalapozott) – ellentétben a korábban említett objektív valószínűséggel –, de mivel feltétel, hogy ezek is teljesítsék a valószínűség-számítás axiómáit, ezért a későbbiekben (formálisan) ugyanúgy lehet velük dolgozni, mint az objektív valószínűségekkel. (A fogadóirodák ezt nap mint nap meg is teszik.)

³ Ez a valószínűség kicsi ugyan, de nem célszerű 0-nak tekinteni. Egyrészt azért, mert ez esetben az egész feladat triviálisan értelmetlenné válik, másrészt meg azért, mert a paranormális képességek nem tartoznak a lehetetlen események közé: egyes országokban komoly kutatóintézetek foglalkoznak ezzel a jelenséggel.

Természetesen a hagyományos módszertan és a bayesi megközelítés közötti vita egyik központi eleme az objektív/szubjektív valószínűség. A bayesi megközelítést bírálják azért, mert megengedhetetlen szubjektivitást visz a döntési, következtetési folyamatba, a bayesi szemlélet hívei pedig az objektív valószínűség hátulütőit hangsúlyozzák. Elsősorban azt, hogy az csak tömegjelenségek esetében értelmes, azaz akkor, ha valóban klasszikus kísérleti elrendezésű feladatokra alkalmazzuk a módszertant. A relatív gyakoriságokon alapuló valószínűség használata tehát csak akkor indokolt, ha e mögött valóban nagy, ismételhető minták állnak. Ez azonban a gyakorlatnak egy igen kis szeletére igaz, a bizonytalan környezetben hozandó döntések vagy következtetések esetében nem jellemző.

A szubjektív valószínűség a bayesi keretek között elsősorban a prior megfogalmazásakor merül fel, így nem csoda, hogy a viták középpontjában is ez áll. Az egyik oldalról azt az érvet hangsúlyozzák, hogy a szubjektív priorok nem egyértelműek, mindenki más és más valószínűséget rendelhet saját meggyőződésének kifejezéséhez, így az eredmény (posterior) is lényegileg tetszőleges lehet. A másik oldalról azt emlegetik, hogy ez ugyan igaz, de az egyre gyűlő, halmozódó bizonyítékok ezt a szubjektivitást előbb utóbb kiküszöbölik, valamint a bizonyítékok súlya alatt a következtetések és a döntések valami objektív határhoz konvergálnak. Ennek azonban ellent mond az a tapasztalat – szól ismét a másik oldal –, hogy egészen különböző világnézetek léteznek ma is, holott az emberiség történelme ezen alapkérdésekben már igen sok bizonyítékot halmozott fel. Különösen élesen merülnek fel ezek a kérdések azokban az esetekben, amikor a döntés vagy a következtetés tárgyáról nincsenek előzetes ismereteink: ekkor olyan priort kell választanunk, ami a lehető legjobban fejezi ki tudatlanságunkat (az ilyet hívják neminformatív priornak). Az ilyen neminformatív priorok konstruálásával kapcsolatban, azok semlegességét vagy objektivitását illetően is komoly viták voltak és vannak.

Nem feladatunk egy ilyen rövid ismertetésben állást foglalni ezekben a lényeges kérdésekben, inkább azt hangsúlyozzuk, hogy mindkét megközelítésnek megvannak az előnyös oldalai és a gyenge pontjai. Ennek demonstrálására bemutatunk egy közismert példát, amit jószerivel ki így, ki úgy értelmezhet, de mindenképpen érdemes az eredményeken elgondolkodni.

3. példa

Egy betegség tesztelésekor meglehetősen furcsa eredmények adódtak, amelyek értékelése a bayesi gondolkodás és a priorok szerepének szép példája. Olyan ritka betegség teszteléséről van szó, mely a népességnek csak egytized százalékában van jelen. Korábbi tapasztalatokból ismert, hogy a nevezett teszt 99 százalékban jelzi a betegséget, ha az tényleg fenn áll, de amennyiben az alany nem beteg, a teszt az esetek 5 százalékában mutat pozitív (hibás) eredményt. A teszt hatásossá-

gának vizsgálatok a következő egyszerű formalizálást végezzük el: legyen A esemény az, hogy a vizsgált személy beteg, B pedig a teszt pozitív eredményét jelölje. Ekkor

$$\Pr(A) = 0,001, \quad \Pr(B|A) = 0,99 \text{ és}$$

$$\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A}) = 0,05094.$$

Bayes tételének alkalmazásával

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} = \frac{0,99 \cdot 0,001}{0,05094} = 0,019.$$

Ez azt jelenti, hogy ha a teszt eredménye pozitív, csak nagyon kicsi a valószínűsége annak, hogy a vizsgált személy valóban beteg. Vagy másként fogalmazva, annak a valószínűsége, hogy a teszt pozitív, de a páciens valójában nem beteg $1 - 0,019 = 0,98$. Ez bizony azt jelenti, hogy az esetek 98 százalékában hamis a riasztás, azaz a teszt igen rosszul működik!

Az eredmény meglehetősen meglehetősen, főleg annak ismeretében, hogy a kiinduló feltételek jó eredménnyel kecsegtettek. Ezért vizsgáljuk meg egy kicsit alaposabban ezt az esetet! Néhány további megállapítás:

1. Ha minden teszt nélkül egy véletlenszerűen kiválasztott páciensről azt tudjuk, annak valószínűsége, hogy beteg 0,001, a pozitív teszt után viszont már ez a valószínűség 0,019-re nőtt, akkor azt mondhatjuk: a teszt tizenkilencszeresre emelte az induló valószínűséget, ami egy egyszerű, tömeges szűrővizsgálat esetén nem is olyan rossz eredmény.

2. Mint láttuk, a hamis pozitív teszt valószínűsége igen nagy, de a teszt teljes értékeléséhez érdemes megnézni a másik oldalt is: mennyi a valószínűsége a hamis negatív tesztnek, azaz mi a valószínűsége annak, hogy valakinél a teszt negatív volt, mégis beteg? A részletek mellőzésével ez a valószínűség 0,0000105, azaz nagyon kicsi, erről az oldalról a teszt hatásosan működik.

3. Végül próbáljuk megfejtetni a hamis pozitív teszt okát. Nem nehéz rájönni, hogy ez a szélsőségesen kis prior valószínűségben rejlik. A prior annyira szélsőséges (távol áll a semleges egyenletes eloszlástól, azaz a $\Pr(A) = 0,5$ -től), hogy döntően befolyásolja a posterior valószínűséget. Ezt könnyű belátni, például úgy, hogy egy gyakori betegségre hasonló peremfeltételeket alkalmazunk; ekkor a teszt már nagyon hatásosan működik.

Ebben a példában a prior adott, nem valami szubjektív választás eredménye. Mégis figyelmeztet arra, hogy szélsőséges esetben milyen anomáliák adódhatnak. Ez nagyon megszívlelendő akkor, amikor az elemző kezében van a prior, azt szubjektív vélekedése alapján készíti el. Akkor bizony nagy a valószínűsége annak, hogy alkalmas (vagy éppen alkalmatlan) priorválasztással tetszőleges, az elemzőnek tetsző eredmény keverhető ki. Ennek kivédésére a cikk végén még visszatérünk.

3. A statisztikai következtetésemélet bayesi keretei

Az eddigiekben általánosságban szóltunk az elemzés, a döntés és az előrejelzés bayesi gondolatvilágáról. Eddigi példáink is ezt mutatták: a bayesi megközelítést egyszerű döntési szituációkon vezettük be, és igyekeztünk olyan általánosan megfogalmazni a mondanivalót, hogy abba lényegileg minden feladat beleférjen. Ezért is lehetett olyan, kissé talán túlzónak tűnő megfogalmazásokat tenni, mint amilyen például az „általános tudományos módszer” volt. A továbbiakban leszűkítjük a vizsgálat körét, és azt nézzük meg közelebbről, kissé konkrétan, hogy a statisztikai következtetéseméletben hogyan, milyen sajátosságokkal jelenik meg a bayesi módszertan.

Kiinduló pontunk itt is Bayes tétele, amit azonban most nem egyes valószínűségekre, hanem egész sűrűségfüggvényekre írunk fel. (Megjegyzendő, hogy a diszkrét eset hasonlóan vizsgálható, de úgy gondoljuk, az általánosabb, folytonos eset elemzése jobban megmutatja a módszertan jellegzetességeit.) A két mennyiség, amelyekre a tételt felírjuk ekkor az y -nal jelölt minta és a θ -val jelölt paraméter lesz:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \propto f(y|\theta)f(\theta), \quad /3/$$

ahol a bayesi elemzésben gyakori \propto jel az arányosságra utal, azaz azt mondja ki, hogy a két oldal egy konstans szorzótól eltekintve egyenlő.

A /3/ összefüggésben θ paraméter jelenti a vizsgálat tárgyát, y minta pedig a tapasztalati eredményeket. Az $f(\theta)$ fejezi ki a paraméterre vonatkozó a priori ismereteinket (feltevéseinket), $f(y)$ az adott minta előfordulását jellemző sűrűségfüggvény, míg igazán a két feltételes valószínűség az érdekes. Ezek közül az $f(y|\theta)$ nem más, mint a közismert likelihood függvény, míg az $f(\theta|y)$ az elemzés vég-eredményét jelentő a posteriori sűrűségfüggvény (posterior), ami azt mutatja, hogy a minta feltételezésével (ismeretében) milyen a vizsgálat tárgyát képező paraméter eloszlása. Ahhoz azonban, hogy ezt, és főleg a /3/ jobb oldalán megjelenő arányosságot megértsük, számba kell vennünk azokat a kereteket, amelyek között a bayesi következtetésemélet érvényes. Ezek ugyanis lényegesen különböznek a klasszikus statisztikában megszokottaktól.

A kétféle szemlélet közötti leglényegesebb különbségeket a következő táblázatban foglaljuk össze.

A klasszikus és a bayesi következtetés lényeges eltérései

Tulajdonság	Klasszikus statisztika	Bayesi statisztika
1. Paraméter	Rögzített	Valószínűségi változó
2. Valószínűség	Objektív	Szubjektív is lehet
3. Külső információ	Nincs vagy csak kevés	Van és lényeges
4. Minta	Valójában csak egy van, de feltételezzük az ismételt mintavétel lehetőségét	Csak egyetlen mintát értékelünk

A táblázat természetesen nem kevés magyarázatot igényel. Az első különbség, ami a két szemlélet között megjelenik, az az, hogy míg a klasszikus statisztikában a becslés és a hipotézisvizsgálat tárgya jellemzően valami ismeretlen, de feltételezés szerint rögzített érték, a bayesi statisztika ezt egy ismeretlen valószínűségi változóként fogja fel. Példaként említhetünk egy egyszerű aránybecslést: legyen becslésünk tárgya a 2010. év egy jól meghatározott időpontjában (például december 31-én 24 órakor) a létminimum alatt élő háztartások aránya. A hagyományos (klasszikus) felfogás szerint ez egy jól meghatározott (bár nem ismert) arány, amelyet – ha megfelelő mérési eszközeink lennének – tetszőleges pontossággal meg lehetne határozni. (Ezt általában azért nem tudjuk megtenni, mert számításaink és becsléseink mintavételi és nemmintavételi hibát tartalmaznak.) A bayesi felfogásban viszont ez egy valószínűségi változó, hiszen értéke sok tényezőtől (külső hatások, gazdaságpolitikai döntések, monetáris és fiskális eszközök alkalmazása stb.) függ. Az, hogy 2010-ben konkrétan egy meghatározott értéket vett fel, nem jelenti azt, hogy ez egy rögzített érték, hanem úgy kell felfogni, mint egy valószínűségi változó egyetlen realizációját. Mindkét felfogás önmagában védhető, ugyanakkor lényeges szemléleti különbség van közöttük. Talán még eklatánsabb példa lehet az előbbiekre valamilyen jól meghatározott makrogazdasági paraméter (például a magyar háztartások valamely időszakra jellemző fogyasztói határhajlandósága), mely esetében talán még inkább érthető, hogy a hagyományos felfogás egy kitüntetett értéket kíván meghatározni, a bayesi pedig a jellemző eloszlását.

A táblázat második pontjával korábban már foglalkoztunk: a klasszikus statisztika kizárólag az objektív valószínűséget fogadja el, míg a bayesi megengedi a szubjektívet is. Megjegyezzük, ezzel nem zárja teljesen ki az objektív valószínűséget, sőt gyakran használja is (likelihood függvény). A vélemények ezzel kapcsolatban azon a ponton csapnak össze, hogy megengedhető-e egyáltalán a szubjektív valószínűség használata, és főleg, keverése az objektív elemekkel.

A harmadik pont ismét nagyon lényeges: ez a külső információ szerepének eltérő felfogása. Külsőnek nevezünk minden mintán kívüli információt. A klasszikus statisztika is használ mintán kívüli információt: elegendő arra gondolni, hogy gyakran

alkalmazunk eloszlásbeli feltételezéseket (amelyeket nem mindig tudunk, olykor nem is akarunk ellenőrizni), vagy például rétegzett mintából történő becslés esetén a rétegek képzése is mintán kívüli információkat használ fel. Ezek az információk azonban mind arra szolgálnak, hogy a mintát javítsák, hozzáférhetőbbé tegyék, particionálják stb. A bayesi esetben viszont intézményesen használunk a mintától független információt annak érdekében, hogy minden olyan körülményt figyelembe tudjunk venni, amely az elemző ismeretkörbe tartozik, de nem testesül(het) meg a mintában.

Végül a negyedik eltérés szintén nagyon lényeges, hiszen a klasszikus statisztika egyik nagy ellentmondására mutat: csak egyetlen minta választására van ugyan módunk, de az ebből levonható következtetéseinket úgy fogalmazzuk meg, hogy feltételezzük az ismételt mintavétel lehetőségét, sőt realitását. Elegendő arra gondolni, ha azt mondjuk egy becslőfüggvényről, hogy az torzítatlan, akkor ezzel hallgatólagosan feltételezzük, az összes lehetséges mintát ki tudjuk választani, hiszen a torzítatlanság csak ezek együttesére állít valamit. Hasonló probléma merül fel a tesztek szignifikanciájánál is: egy szignifikáns teszteredmény csupán annyit jelent, hogy ha módunk lenne nagyon sokszor ismételt elvégezni az eljárást, akkor... De a valóságban erre soha nincs módunk. Mindezzel szemben a bayesi statisztika számol azazal, hogy csak egyetlen mintánk van, a következtetéseket abból kell levonnunk, és csak annak keretében lehet/szabad értelmeznünk. Ismételt mintavétel itt tehát szóba sem jön. Ennek ismeretében kell értelmezni a /3/ összefüggés jobb oldalát: mivel csak egyetlen minta kiválasztásában gondolkodunk, az ehhez a mintához tartozó sűrűségfüggvény csupán egy (konstans) értéket határoz meg, ami valójában a teljes valószínűség tételéből vezethető le. Ekkor pedig felírhatjuk az utolsó arányossági relációt, ami konstans szorzótól eltekintve egyenlőséget jelent, és megfogalmazva annyit tesz, hogy *a posterior arányos a likelihood és a prior szorzatával*. Ez a bayesi következtetéselmélet alapja és kiindulópontja.

Az, hogy ezt az arányosságot hogyan változtathatjuk egyenlőséggé (ha szükséges), már csak technikai kérdés. Két lehetőség adódik: az egyik az, hogy az $f(y)$ sűrűségfüggvényt a teljes valószínűség tétele formájában kifejezzük:

$$f(y) = \int_{\theta} f(y|\theta)f(\theta)d\theta.$$

A másik, gyakrabban alkalmazott, és a módszer szellemével inkább összhangban álló megoldásként kihasználjuk azt, hogy a posteriornak is sűrűségfüggvénynek kell lennie, azaz integráltja megfelelő tartományon 1. Ekkor a /3/ jobb oldalán megjelenő függvényhez olyan konstans szorzót rendelünk, amely biztosítja, hogy a posterior is sűrűségfüggvény legyen. (Ezt az eljárást szoktuk *kiintegrálásnak* nevezni.) Megjegyezzük, hogy ez az elvben egyszerű művelet a gyakorlatban, főként bonyolultabb,

sokdimenziós feladatok esetén komoly nehézségekbe ütközik, amelyekről e cikk befejező részében még szólunk.

Az elmondottakból már látható, hogy a bayesi felfogás valóban fenekestől felfordítja a statisztikai következtetéseméletet: olyan biztos alapok szűnnek meg, illetve válnak kérdésessé, amelyeket a klasszikus statisztikában külön nem, vagy csak alig kell említeni, értelmezni, és olyan alapelvek születnek, amelyekre a klasszikus statisztikán nevelkedettek nem is gondolnak.

Folytatva a következtetésemélet tárgyalását, a /3/ alapegyenlet egyes elemeivel célszerű egy kicsit behatóbban is foglalkozni. A likelihood függvényt nem kell részletezni; azt minden haladó statisztikai kurzuson alaposan tárgyalják, megmutatják tulajdonságait, és részletesen foglalkoznak a likelihood függvényre épülő maximum likelihood (ML) becslésekkel is. A likelihood függvényt a minta tulajdonságai, illetőleg a kapcsolódó feltevések egyértelműen meghatározzák, így az mintegy technikai adottság az elemző részére. Nem ez a helyzet a prior eloszlással, illetve sűrűségfüggvénnyel kapcsolatban. A prior ugyanis nem automatikusan adódik, hanem előzetes ismereteinket, vélekedésünket, állításainkat kifejező függvény. Aligha kell hangsúlyoznunk, hogy a prior konstruálásában az elemzőnek nagy szabadsága van, hiszen szubjektív vélekedését sok különböző formában ki lehet fejezni. A prior konstruálásakor valójában három lényeges pontot kell szem előtt tartani. Az első az, hogy a prior legyen független a mintától, azaz olyan információkat tartalmazzon, amelyek nem testesülnek meg a mintában. Ezt persze nem mindig könnyű biztosítani, ám az is azonnal belátható, hogy ha a prior a mintát vagy részben a mintát ismétli meg, akkor az elemzés értéke ezáltal romlik. Még akkor is, ha ennek jellemzésére, mérésére nincs alkalmas mutatószámunk. Azonban, ha a priort időben a mintavétel előtt felállítjuk, jó esély van a két információforrás függetlenségére. A másik követelmény, amit célszerű a prior készítésekor szem előtt tartani, az, hogy olyan függvényformát válasszunk, amelyik a paraméter(ek) mozgásával, finomhangolásával különböző induló feltételezéseket képes leírni. Természetesen, a több paramétert tartalmazó eloszlások esetén erre több lehetőség adódik, de gyakran már két paraméterrel is sok különböző alakú, jellemzőjű sűrűségfüggvényt tudunk képezni. Végül a prior megválasztásakor célszerű bizonyos kényelmi, kezelhetőségi szempontokat is figyelembe venni: ha van választásunk (és többnyire van), akkor törekedjünk arra, hogy az így választott prior és a feladat által determinált likelihood szorzata lehetőleg kényelmesen kezelhető legyen, sőt adjon lehetőséget a továbblépésre (láncolásra). Abban a speciális esetben, amikor a posterior típusát illetően megegyezik a priorral, azaz egy esetleges következő lépésben priorként szolgálhat, azt mondjuk, hogy a prior *természetes konjugált*. Bonyolult modellek estén az efféle kényelmes megoldást jelentő konjugált priorok nem jellemzők. Részben ez az oka annak, hogy az utóbbi években olyan gyorsan fejlődtek és terjedtek el a különféle számítógépes eljárások ilyen feladatok megoldására.

4. példa

Aránybecslési feladatok esetében a likelihood függvény magja $P^k(1-P)^{n-k}$ alakú. Ha ehhez egy béta-eloszlású priort választunk, akkor a posterior is béta-eloszlású lesz, azaz természetes konjugált, hiszen a béta-eloszlás sűrűségfüggvénye

$$f(x) = kx^{\alpha-1}(1-x)^{\beta-1},$$

ahol α és β paraméterek, k pedig a paraméterektől függő, de általuk pontosan meghatározott normalizáló konstans. Ha priorként ezt az aránybecslés ismeretlen P paraméterére alkalmazzuk (ezzel fejezve ki előzetes ismeretünket a paraméterre vonatkozóan), akkor látható, hogy a likelihood, a prior és szorzatuk eredményeképpen a posterior is azonos típusú eloszlás lesz.

A prior eloszlás típusa mellett – természetesen – paramétereit is meg kell határozni; ebben testesülnek meg az a priori információk. Itt gyakran használják a momentumok módszerét, hiszen egy eloszlás momentumai (például várható érték, variancia) olyan jellemzők, amelyekhez viszonylag könnyű köznapi tartalmat kapcsolni, így az előzetes, szubjektív információkat jól lehet ezeken keresztül bevezetni a prior eloszlásba, azon keresztül pedig a modellbe.

A priorok konstruálásakor fontos kérdés az, hogy mi történjék akkor, amikor nincs semmiféle előzetes feltevésünk a vizsgálat tárgyáról. A módszertan zártsága azt követeli meg, hogy az eljárás (posterior konstruálása) erre az esetre is működjön, azaz olyan priort kell találnunk, amelyik kifejezi a teljes információhiányunkat. Talán furcsának tűnik, de ez a feladat – a neminformatív priorok készítése – egyáltalán nem könnyű, nagyon sok vitát váltott és vált ki a teoretikusok, valamint alkalmazók körében egyaránt. Egyszerűbb esetekben az az elv vált általánossá, hogy a tudatlanságot az entrópia maximalizálás útján fejezzük ki, azaz olyan priort keresünk (milyen halmazból?), amelyiknél az eloszlás entrópiája a lehető legnagyobb. Egyszerű esetben (például, amikor a 4. példában említett P sokasági arány (valószínűség) a vizsgálat tárgya), a $(0,1)$ intervallumban értelmezett egyenletes eloszlás (ami nem más, mint egy béta $(1,1)$ eloszlás) felel meg ennek a követelménynek, de rögtön bonyolultabbá válik a helyzet akkor, ha például nem véges az az intervallum, amelyen a keresett paraméter értékeket vehet fel.

A priorok készítésének speciális esete az, amikor nem szubjektív meggyőződésünk, hanem egy korábbi mintán alapuló felvétel alapján szeretnénk priort készíteni. (Más szóval, amikor a bayesi módszertant, mint eszközt több különböző kutatás eredményének összekapcsolására használjuk. Ilyen feladatok nem ritkán adódnak a manapság bizonyos szakmákban meglehetősen népszerű metaanalízis kapcsán.) Az ilyen priorokat objektív, adatokon alapuló (data-based) prioroknak nevezzük.

Az eddigiekből jól látható, hogy a priorok készítése nem mechanikus feladat, nagyfokú tárgyismeretet, technikai tudást és nem kevés kreativitást feltételez. Ezzel szemben a posterior előállítását innen már lényegileg semmi érdemi feladatot nem jelent: egyszerű szorzással előállítható a posterior magja, jóllehet ez esetenként sok technikai problémát okozhat, amelyekről később még szólnunk. Valamit azonban még mondanunk kell a posterior értelmezéséről. A posterior (megfelelő normálás után) sűrűségfüggvény együttesen tartalmazza az előzetes, mintán kívüli, valamint a mintában megtestesülő információkat. Amennyiben egydimenziós feladatról van szó, azt is mondhatjuk, hogy összesúlyozza a kétféle módon kapott mennyiségeket, és az eredmény odahúzza, ahol biztosabb, kisebb szórású információk találhatók. Ha előzetes tudásunkat nagyon biztosnak tartjuk, kisebb szórást rendelünk hozzá (szubjektív módon), a prior jellemzői fognak dominálni a posteriorban és viszont. Sőt szélsőséges esetben, ha előzetes ismereteinket teljesen biztosnak tekintjük, a posteriorban azt is kapjuk vissza – a mintának a tapasztalatoknak így semmi teret nem hagyva. Amennyiben neminformatív priorból indulunk ki, azt várhatjuk – és az esetek jó részében ez így is van –, hogy a posteriorban a likelihood ismétlődik meg: ez esetben – igaz más szemléletben –, de visszakapjuk azokat az eredményeket, amelyeket a klasszikus statisztikából ismerünk. Azonban két dolgot ekkor se felejtünk el! Egyrészt az összefüggések általában bonyolultabbak annál, mintsem, hogy a posteriort ilyen egyszerűen, súlyozott átlagként fogjuk fel. Másrészt, még ha formálisan a klasszikus eredményekhez hasonlókat kapunk is, tartalmilag emlékezzünk arra (lásd a táblázatot), hogy egészen más szemlélet és háttér áll a bayesi modell mögött. Ezért kell óvatosan kezelni az olyan megállapításokat, miszerint a nagymintás bayesi statisztika a likelihoodhoz tart. Igaz ugyan, hogy rögzített prior esetén, ha a minta méretét növeljük, a posteriorban egyre inkább a nagy minta válik dominánssá, amire például a ML nagymintás eredményeit lehet építeni, de ne feledjük, hogy ezek az eredmények mást jelentenek, mint a hagyományos szemléletben.

4. Bayesi becslés és hipotézisvizsgálat

A bayesi elemzésnek a posterior elkészítésével valójában vége van: a posterior a bayesi következtetéselmélet végeredménye, hiszen ez a függvény minden információt tartalmaz a keresett paraméterre. Az elemzés mégsem áll meg itt, hanem folytatódik, egyrészt azzal a céllal, hogy jobban értelmezhető, kezelhető végeredményeket kapjunk, emellett olyan mutatókat tudjunk származtatni, amelyek megfelelnek a megszokott klasszikus fogalmaknak, ezáltal összehasonlíthatóvá téve a két elemzési utat. De lényeges ága az elemzésnek a hipotézisvizsgálat is, amely bayesi keretek között egészen más, mint hagyományos kontextusban.

4.1. Intervallum- és pontbecslés

Bár a bayesi elemzés eredményének a posterior sűrűségfüggvény felel meg, ennek jobb áttekintése érdekében a benne rejlő információkat olykor célszerű sűríteni. Így jutunk el az intervallum- és pontbecslés feladatához. A sorrend nem véletlen, mert bár a klasszikus statisztika előbb tárgyalja a pontbecslést, aztán a pont köré képez intervallumot, ez esetben elsődleges az intervallum, hiszen a posterior sűrűségfüggvény készen áll. Az intervallumbecslés feladatával kapcsolatban két fontos mozzanatot kell kiemelni. Az egyik az, hogy a klasszikus statisztikában megszokott szimmetrikus intervallum helyett célszerű a *legsűkebb intervallum* fogalmát bevezetni. Ha egyszer megadunk egy lefedettség mutatószámot (mondjuk 95 százalékot), akkor általános esetben végtelen sok olyan, egymástól különböző intervallum készíthető, amelyek esetén a posterior eloszlás sűrűségfüggvénye alatti terület adott (itt 95 százalék). Ezek közül kitüntetjük azt, amelyik a legsűkebb, azaz, amelyik a lehető „legpontosabb” becslést eredményezi. Ezt az intervallumot általában csak gépi úton, kereséssel lehet meghatározni. Az intervallumbecsléssel kapcsolatos másik fontos megjegyzés annak tartalma, értelmezése. A klasszikus statisztikában megszokott értelmezés itt (bayesi statisztika) nem helytálló, hiszen ott (klasszikus statisztika) úgy fogalmazunk, hogy „a minták 95 százalékára igaz, hogy az így számított intervallum tartalmazza a keresett jellemző paramétert”. Ez esetben, mivel csak egyetlen mintánk van, „mindössze” olyan értelmezést adhatunk az intervallumnak, hogy „a jellemző, paraméter lehetséges értékeinek 95 százalékát tartalmazza”. A százalék ekkor úgy értendő, ahogy azt a priorban értelmeztük.

A pontbecslés feladata úgy fogalmazható meg, hogy keressük azt a pontot (jellemzőt, paraméterértéket), amely valamilyen szempontból a lehető legjobban jellemzi a posterior eloszlást. A feladat nem egyértelmű, hiszen ahhoz, hogy ezt a kitüntetett pontot megtaláljuk, pontosan meg kellene fogalmazni a vele szemben támasztott követelményeket. Általában az a szokás, hogy valamilyen veszteségfüggvényt definiálunk, ami kifejezi, ha nem az adott pontot választanánk az egész eloszlás jellemzésére, akkor súlyozottan mennyi információt vesztenénk. Például, ha a sűrűségfüggvény szélén jelöljük ki ezt a kitüntetett pontot, akkor a magas függvényértékű (és így nagy súlyú) értékek messze kerülhetnek tőle, ami növeli a veszteséget. Célszerű veszteségfüggvénynek valami szimmetrikus függvényt választani, ami a mindkét irányú eltéréseket azonos módon bünteti, a nagy eltérést természetesen jobban. Ha a posterior változót θ -val, a kitüntetett pontot (pontbecslést) $\hat{\theta}$ -pal jelöljük, akkor vagy a $(\theta - \hat{\theta})^2$ kvadratikusan vagy a $|\theta - \hat{\theta}|$ abszolútérték-függvényt tekintjük veszteségfüggvénynek, és azt a $\hat{\theta}$ értéket tekintjük pontbecslésnek, amelyikre a veszteségfüggvény várható értéke minimális. Egyszerű levezetéssel igazolható, hogy

a kvadratikus veszteségfüggvény a posterior várható értéket, az abszolút értékre épülő veszteségfüggvény a posterior mediánt eredményezi pontbecslés gyanánt. A veszteségfüggvények esetén további kényelmes választás az ún. „mindent vagy semmit” veszteségfüggvény, amely esetén a pontbecslés a posterior módusz lesz, illetve súlyozott abszolútérték-veszteségfüggvény esetén a pontbecslés a súlyok által meghatározott kvantilis lesz.

Ha ezek közül bármelyiket is választjuk pontbecslésnek (a várható érték választása a leggyakoribb), azzal jellemezhetjük a posterior eloszlást, és összevethetjük azt a klasszikus becslés során kapott megfelelő paraméterrel. Soha ne feledjük azonban, hogy ez az összevetés a kiinduló feltételek eltérő volta miatt csak korlátozott érvényű, hisz mindkét becslés csak a saját feltételrendszere között értelmezhető. Arra is fel kell hívni a figyelmet, hogy a bayesi pontbecslés esetén értelmetlen olyan tulajdonságokat keresni, amelyek csak ismételt mintavételnél érvényesek (például torzítatlanság, konzisztencia stb.). Végül megemlítjük, hogy a bonyolult valós modellek esetében a posterior jellemzők numerikus meghatározása nem egyszerű feladat, ezért megoldása komoly számítástechnikai háttérrel tétel fel.

4.2. Hipotézisvizsgálat

A bayesi hipotézisvizsgálat – hasonlóan a becsléshez – lényegesen eltér a klasszikus statisztikában megszokottól. Bár kiindulópontjuk, a nullhipotézis, valamint az ellenhipotézis fogalma és megfogalmazása látszólag hasonló, valójában már itt is lényegi szemléletbeli különbség van. Míg ugyanis a klasszikus statisztikában e két hipotézist nem kezeljük szimmetrikusan, a bayesi statisztika hipotézisei szimmetrikusak és megfordíthatók. A klasszikus statisztikában a nullhipotézis kitüntetett szerepet játszik, és döntésünk valójában nem az, hogy a nullhipotézis vagy az ellenhipotézis igaz, hanem a nullhipotézis elutasítható/elutasítandó-e, avagy ennek híján el kell-e azt fogadnunk. Figyeljük meg, ez más, mint az egyszerű választás a két hipotézis között!

A bayesi teszteljárás két kulcsfogalma (a hipotézisek mellett) a prior és a posterior esélyhányados (odds), melyek definíciói:

$$PRO = \frac{\Pr(H_1)}{\Pr(H_0)}, \text{ illetve } POO = \frac{\Pr(H_1|y)}{\Pr(H_0|y)}.$$

A prior esélyhányados (*PRO*) azt mutatja meg, hogy induláskor, mintavétel előtt melyik hipotézist tartjuk valószínűbbnek, míg a posterior esélyhányados (*POO*) ezt a mintavétel után, a mintában rejlő információk feldolgozását követően méri. A két

mennyiség közötti összefüggés egyetlen θ paraméter esetén (amelyikre a hipotézisek vonatkoznak) folytonos esetben:

$$POO = \frac{\Pr(H_1)}{\Pr(H_0)} \cdot \frac{\int \Pr(y|\theta, H_1) \Pr(\theta|H_1) \partial\theta}{\int \Pr(y|\theta, H_0) \Pr(\theta|H_0) \partial\theta},$$

ami nem más, mint a prior esélyhányados és egy átlagos likelihood hányados szorzata. Ez a második tényező azt mutatja, hogy miként változott a két hipotézis megítélése a mintavétel következtében. Ez valójában a Bayes-faktor legegyszerűbb változata, és ez az alapja a bayesi szemléletű információs kritériumoknak, amelyeket a modellválasztásban kiterjedten használnak.

A hipotézisvizsgálat ezek ismeretében már egyszerű: ha a posterior esélyhányados egynél nagyobb, akkor mintavétel után az ellenhipotézist tartjuk esélyesebbnek, és azt fogadjuk el, míg ha egynél kisebb, akkor a nullhipotézist. A mutatók számítása viszonylag bonyolult is lehet, de maga a döntés egyszerű. Egyrészt figyeljük meg azt, hogy a két hipotézist (amelyek egyébként itt is egymást kizárók, mint a klasszikus esetben) teljesen szimmetrikusan kezeljük. Másrészt tudomásul kell vennünk, hogy ez esetben a klasszikus teszteknel megszokott olyan fogalmak, mint a szignifikanciaszint, vagy a p -érték, értelmetlenek, hiszen azok a minták összességére tesznek valamiféle megbízhatósági állítást. Ebben az esetben ez haszontalan és felesleges. Valójában egy egyszerű, szinte leíró mutató értéke alapján hozzuk meg a döntésünket.

5. Technikák és alkalmazások

Miután nagyon röviden áttekintettük a bayesi gondolkodás és közelebbről a statisztikai következtetések lényegi kérdéseit, szólnunk kell néhány szót a számítások, modellek megvalósításáról is. Már az eddigiek alapján is látható, hogy a bayesi elemzések során a klasszikus statisztikában megszokott technikák helyett más irányú számításokon van a hangsúly. A bayesi statisztika jellemző számításigényes fázisai a posterior értékeléséhez és az az alapján történő elemzésekhez kapcsolódnak. Említettük a konstans kiintegrálását, de idetartozik a sokdimenziós eloszlások peremeinek meghatározása, bonyolult eloszlások várható értékének, mediánjának, momentumainak meghatározása, a legszűkebb intervallum számítása és a posterior esélyhányadosok kiszámítása. Ezeknek a számítási feladatoknak a nagy része a függvények numerikus kezelésére (integrálás, szélsőérték-számítás) vezethető vissza. Részben ennek

az igénynek a hatására alakultak ki az utóbbi évtizedekben azok az algoritmusok, amelyek ezeket a feladatokat különféle mintavételeken alapuló szimulációs technikákkal oldják meg. Kiemelkedő közülük a nagyon hatékonyak bizonyuló és kivételes népszerűsége szert tevő Markov-láncokon alapuló Monte-Carlo-algoritmusok (Markov Chain Monte Carlo – MCMC), amelyeket egyesek minden idők tíz legfontosabb algoritmusai egyikeként aposztrofálnak (*Andrieu et al.* [2003], *Geweke* [2005]). Ezek lényege: a vizsgálni kívánt eloszlásokból mintát vesznek úgy, hogy egy alkalmasan kidolgozott Markov-lánc vezeti a mintavételt. A minták empirikus jellemzőiből pedig következtetni lehet azok analitikusan nem, vagy csak nehezen vizsgálható tulajdonságaira. A MCMC jelenleg is a kutatások homlokterében áll, de sok algoritmus már forgalomban levő szoftverekben (például R, WinBUGS, OpenBUGS) elérhető. Az MCMC-módszerről magyar nyelven *Kovács–Balogh* [2009] cikkében olvashatunk.

Ami a bayesi statisztika alkalmazásait illeti, azok igen kiterjedtek. Teljesen reménytelen vállalkozás lenne még ennél jóval nagyobb lélegzetű műben is kísérletet tenni ezek valamennyire is teljes vagy akár reprezentatív áttekintésére. Ezért itt csak teljesen szubjektíve emelünk ki néhány jellemző példát. Érdekes megemlíteni, hogy az alkalmazások gyakran a bayesi statisztikának inkább csak a technikáját veszik át, a valódi mögöttes keretrendszert nem. Így olykor furcsa, öszvérmegoldások születnek. Csupán a társadalmi-gazdasági elemzés területéről hozva a példákat (hiszen más szakterület eredményeire a szerzőnek nincs elegendő rálátása), kezdetben gyakoriak voltak az egyszerű becslésekhez (arány, átlag, értékösszeg) kapcsolódó külső információs becslések. Az ökonometriában az idősoros modellek paramétereire vonatkozó keresztmetszeti információkat egyebek között bayesi keretekben kezelték (*Zellner* [1978]). Nagy hatású volt *Shiller* [1971] tanulmánya, melyben az osztott késleltetésű modellek becslési problémáira talált rugalmas bayesi megoldást. Ugyanerre az ötletre építve mutatható meg például a ridge-regresszió néhány fontos tulajdonsága. Újabban a hagyományos elemzési feladatokban megszokottnál nagyobb információs halmazra épülő, sok ismeretlen paramétert tartalmazó becslések esetén alkalmazzák kiterjedten a bayesi módszereket. Az ilyen feladatok közül kiemelendők a BVAR- (Bayesian vector autoregressive) és a látens változós modellek, a dinamikus faktormodellek, valamint általában azok, amelyeknél a nagyszámú paraméter önmagában nem lenne identifikálható.

A magyar gazdaságmodellezési gyakorlatban sajnos nagyon kevés bayesi indítással írással lehet találkozni. Ezek közül talán *Gál* [1998], *Várpalotai* [2008], valamint *Kovács* és *Balogh* [2009] munkáit kell megemlíteni. Gál egy biztosító saját állományára vonatkozó halálozási valószínűségeit korévenként becsülte. A priort országos halandósági táblákból vette át, és béta-eloszlást feltételezett, míg a minta szerepét a biztosító saját állományának halálozási adatai játszották. Az évenként becsült posterior valószínűségeket mozgó átlagolással simította, és így kapott jól használható

halálzási valószínűségi görbéket. Várpalotai a Shiller-féle DL- (distributed lag) modellt alkalmazta árbegyűrűzések vizsgálatára, valamint ezek előrejelzésére. Modelljeiben ő használt először Magyarországon MCMC-módszereket. Kovács és Balogh a sertésárak alakulását jelezték előre a bayesi statisztika segítségével.

Az utóbbi években a szakirodalom tanúsága szerint számos területen fordulnak a statisztikusok a bayesi módszertan felé, melynek népszerűsítéséből a külön erre a célra alakult tudományos társaság is kiveszi részét (<http://bayesian.org>). Az egyre növekvő és könnyen hozzáférhető információs források (elektronikus adatgyűjtés, internet stb.), valamint a növekvő számú és könnyebben hozzáférhető szoftverek abba az irányba mutatnak, hogy mind több információt lehet beépíteni a modellekbe. Fontos azonban, hogy ennek ne csak módszertanával, de az alkalmazás feltételrendszerével és az eredmények helyes értelmezésével is tisztában legyenek az alkalmazók.

6. Összefoglalás

A bayesi gondolkodás azon az egyszerű modellen alapul, hogy meglévő tudásból indul ki, amelyet kombinálva, konfrontálva az adatok által közvetített tapasztalatokkal, a tudás magasabb szintre emelhető. Mindez támpontot nyújt ahhoz, hogy bizonytalan helyzetekben dönthessünk, jó keretet teremt a tudományos gondolkodásnak, alapot ad a különböző szakterületeken a módszer helyes kialakítására. Ugyanakkor a bayesi megközelítés egy sor elemét (elsősorban a szubjektív valószínűség létjogosultságát a tudományos módszertanban, és a szubjektív priorokat) máig is erősen vitatják. Így van ez a statisztikában is, amelyet talán a leginkább közvetlen módon érintett a bayesi gondolkodás. A bayesi statisztika ennél fogva gyökeresen különbözik a klasszikustól: mások az induló feltevései, mások az eszközei és természetesen eredményei is. Igen lényeges különbség az, hogy a bayesi statisztika szakít a klasszikus statisztika ismételt mintákon alapuló koncepciójával, és ebből következően minden ezekhez kapcsolódó eredményt (becslési kritériumok, konfidenciasávok, szignifikanciaszintek stb.) negligál, illetve átértelmez. Ezzel a klasszikus statisztika egy sajátos ellentmondását oldja fel („egy mintából számolunk, több minta feltételezésével értékelünk”). Igaz, helyette bevezeti a már említett szubjektivitást, aminek keveredése az objektív valószínűségekkel komoly értelmezési nehézségeket vet fel. Meg kell jegyeznünk, hogy a szubjektivitás okozta gondokat némiképp enyhítheti, ha az elemző feltevéseit pontosan és transzparens módon fogalmazza meg és mutatja be. Ezzel egyébként helyes utat mutathat a statisztikai etika felé is, hiszen a számítások és az eredmények reprodukálhatósága fontos, de nem mindig teljesülő követelmény.

Nem feladatunk a klasszikus és a bayesi statisztika közötti közel évszázados vitában állást foglalni, hiszen a két irányzat mindegyike értékes eredmények bölcsője, sőt nem ritkán nem egymással szemben állva, hanem egymást erősítve, megtámogatva léteznek. A gyakorlatban rohamosan tör előre a bayesi szemlélet, de az alkalmazott statisztika alapjai még mindig szilárdan klasszikus szemléletűek, és a bayesi módszerek is nagyon gyakran ebből táplálkoznak. A statisztikaoktatásában ugyanakkor a klasszikus szemlélet egyeduralkodó, bár sok helyütt (igaz, inkább doktori képzésben) oktatnak már bayesi szemléletű statisztikát. Megítélésünk szerint a két egymástól nagyon különböző felfogású módszertan tartós együttélésre van ítélve: a feladatok egy részénél az egyik, más részénél a másik kerülhet előtérbe. Ahhoz azonban, hogy ez az együttélés békés, sőt mindkét irányzat számára hasznos legyen, művelőiknek (elsősorban a klasszikus statisztikáénak) egymás eredményeit kellene kölcsönösen és alaposan megismerniük. Ezért az ismeretterjesztés minden szinten fontos: célszerű lenne bayesi bevezetőt adni már egyetemi statisztikai alapkursusokon, a felsőági oktatásban a bayesi kurzusok nagyon kellenének, és a szakmai folyóiratok részéről is hiányzik biztatás a bayesi indíttatású tanulmányok publikálására. Ez a dolgozat is, a maga szerény eszközeivel, ezt a célt szeretné szolgálni.

Irodalom

- ANDRIEU, CH. – DE FREITAS, N. – DOUCET, A. – JORDAN, M. I. [2003]: An Introduction to MCMC for Machine Learning. *Machine Learning*. Vol. 50. No. 1–2. pp. 5–43.
- BAYES, T. [1958]: An Essay towards Solving a Problem in the Doctrine of Chances. 1763. Reprint. *Biometrika*. Vol. 45. No. 3–4. pp. 293–315.
- CANOVA, F. [2007]: *Methods for Applied Macroeconomics Research*. Princeton University Press. Princeton.
- GÁL P. [1998]: *Halandósági táblák becslése bayesi módszerekkel*. Rajk László Szakkollégium. Budapest.
- GEWEKE, J. [2005]: *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc. Hoboken.
- HUNYADI L. [2001]: *Statisztikai következtetésemélet közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- KADANE, J. B. [2009]: Bayesian Thought in Early Detective Stories. *Statistical Science*. Vol. 24. No. 2. pp. 238–243.
- KOOP, G. [2003]: *Bayesian Econometrics*. John Wiley & Sons, Inc. New York.
- KOVÁCS S. – BALOGH P. [2009]: Bayesi statisztikával becsült nem stacionárius idősorok a sertésárak előrejelzésében. *Statisztikai Szemle*. 87. évf. 10–11. sz. 1058–1077. old.
- LINDLEY, D. V. [1965]: *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge University Press. Cambridge.
- SAVAGE, I. R. [1962]: The Subjective Basis of Statistical Practice. In: *Technical Report, Department of Statistics*. University of Michigan. Ann Arbor.

- SHILLER, R. [1973]: A Distributed Lag Estimator Derived from Smoothness Priors. *Econometrica*. Vol. 41. No. 4. pp. 778–788.
- VÁRPALOTAI V. [2008]: *Modern bayesi ökonometriai elemzések. Simasági priorok alkalmazása az üzleti ciklusok szinkronizációjának mérésére és az infláció előrejelzésére*. PhD-értekezés. Budapesti Corvinus Egyetem. Budapest.
- WICKMANN, D. [1995]: *Bayes-statisztika*. ELTE Eötvös Kiadó. Budapest.
- WIKIPEDIA: *Bayesian inference*. http://en.wikipedia.org/wiki/Bayesian_inference
- ZELLNER, A. [1971]: *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, Inc. New York.

Summary

The article summarizes the key elements of Bayesian thinking and particularly those of the Bayesian statistical inference. Its starting point is the original Bayes theorem which can be regarded as a general frame of modelling, moreover, as the basis of the scientific methodology according to the author's demonstration. While discussing the statistical inference, the article highlights the points distinguishing the classical and the Bayesian statistics. Among the main elements of Bayesian inference, the construction of the different priors is underlined. The discussion is focused on the analysis of the posterior, and notably on the posterior indicators (like posterior mean and variance) which can form bridges between the classical and Bayesian results. Bayesian test procedures are touched upon, and the special tools of numerical computations (like MCMC algorithms) are emphasized as well. From the host of applications, some Hungarian economic modelling attempts are quoted. As a final conclusion, the author states that the two approaches are complementary rather than competing but the closer cooperation and further development of both lines requires better understanding of these methodologies. This suggests the responsibility of education and that of the scientific publications.