



BAYES-I MODELLEZÉS A GYAKORLATBAN

TEJELŐ TEHÉNÁLLOMÁNYOK ÁLLOMÁNYON BELÜLI PARATUBERKULÓZIS-ÉRINTETTSÉGÉNEK BECSLÉSE I.

A közlemény a Magyar Állatorvosok Lapja 2024. júniusi számában megjelent cikk másodközlése

Veres Katalin^{1*}
Lang Zsolt¹
Monostori Attila²
Ózsvári László³

¹ ÁTE, Gazdaságtudományi és Biostatistikai Intézet, Biostatistika Tanszék

² Állattenyésztési Teljesítményvizsgáló Kft.

³ ÁTE, Gazdaságtudományi és Biostatistikai Intézet, Törvényszéki Állatorvostani és Gazdaságtudományi Tanszék

ÖSSZEFOGLALÁS

A Bayes-i módszertan széles körben elterjedt fertőző betegségek prevalenciájának modellezésére az állatorvosi irodalomban, hiszen ezzel a megközelítéssel a korábbi ismeretek és az új adatok egyszerre építhetők be a becslésekbe. A szerzők áttekintik a Bayes-i modellezés alapvető fogalmait és működési elveit, továbbá bemutatják, hogyan alkalmazható ez a módszer a gyakorlatban a

paratuberkulózis valódi prevalenciájának telepi szintű becslésére tejelő szarvasmarhák esetében (<https://github.com/VeresKatalin/PTBC>). A modell eredményei alapján egyszer ellett tehenek esetében a valódi prevalencia a látszólagos prevalencia 1,6-szorosaként, többször ellett teheneknél pedig 1,5-szöröseként becsülhető.

Bayes-i modellezés

A Bayes-i módszertan széles körben elterjedt fertőző betegségek prevalenciájának modellezésére az állatorvosi irodalomban. A HANSON és mtsai által javasolt keretrendszer felhasználásával számos, e módszertant alkalmazó tanulmány született az utóbbi évtizedben. A módszer gyors terjedésének fő oka az, hogy a Bayes-i megközelítés segítségével a korábbi ismeretek és az új adatok egyszerre építhetők be a becslésekbe. Fertőző betegségek esetén pl. az országos prevalencia vagy a korábbi

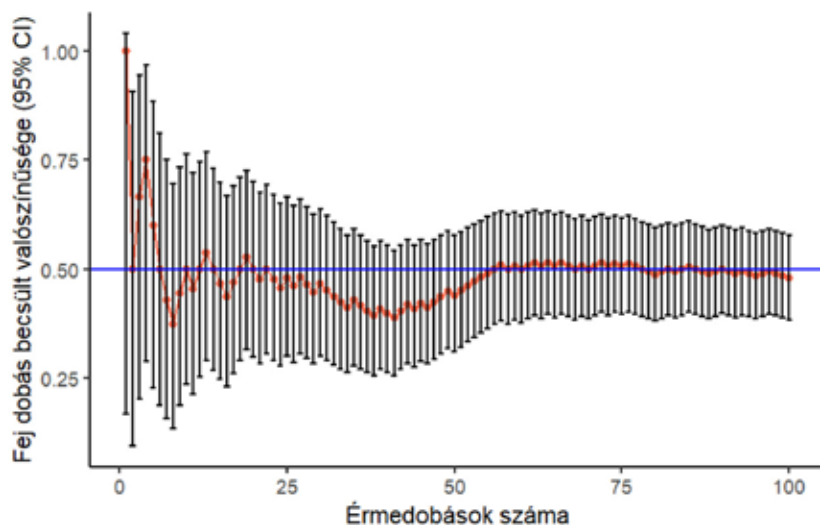
évek felméréseiből származó adatok egyaránt olyan hasznos információt hordoznak, amelyeket hiba lenne figyelmen kívül hagyni. Sokan mégis félve közelítenek ehhez a módszertanhoz, hiszen hagyományosan az ún. frekventista (p -értékekre támaszkodó) statisztikát alkalmazzuk. A közleményünk célja ezért a Bayes-i módszertan legfontosabb fogalmainak, működésének áttekintése és gyakorlati alkalmazásának illusztrálása tejelő szarvasmarha állományok paratuberkulózis- (PTBC) érintettség becslésének bemutatásával.



A Bayes-i statisztika valószínűségfelfogása alapjaiban különbözik a hagyományosan alkalmazott frekventista statisztikától. A frekventista statisztika egy esemény valószínűségét sok kísérlet alapján kiszámított előfordulási részarány, azaz *relatív gyakoriság* alapján határozza meg. Minél többször végezzük el a

kísérletet, annál jobban megközelíti az esemény relatív gyakorisága az elméleti valószínűséget. Pl. minél többször dobunk fel egy szabályos pénzérmét, annál jobban megközelíti a fejek (és az írások) részaránya a várt 50%-ot (**1. ábra**).

1. ábra: Érmédobás szabályos érmével



Piros: fej dobások aránya, fekete: 95%-os konfidenciaintervallum

E keretrendszerben a valószínűség objektív mennyiség, független minden korábbi tapasztalattól. Több egymás utáni fej dobása után továbbra is azt feltételezzük, hogy a következő dobásnál egyaránt 50–50% a fej, ill. az írás dobásának esélye. Egy modell paramétereinek frekventista becslésénél ezt a rögzített valószínűséget szeretnénk meghatározni.

A Bayes-i módszertanban a valószínűség megadásakor az adatok mellett az előzetes tapasztalatokat és ismereteket is figyelembe vesszük. Ez a módszer közel áll hétköznapi gondolkodásunkhoz. Az érmefeldobás példájánál maradva, ha a sokadik dobásra is fejet kapunk, gyanakodni kezdünk, hogy az érme nem szabályos, tapasztalataink alapján „újraszámoljuk” magunkban az esélyeket, és a következő dobásoknál nagyobb eséllyel várunk fejet.

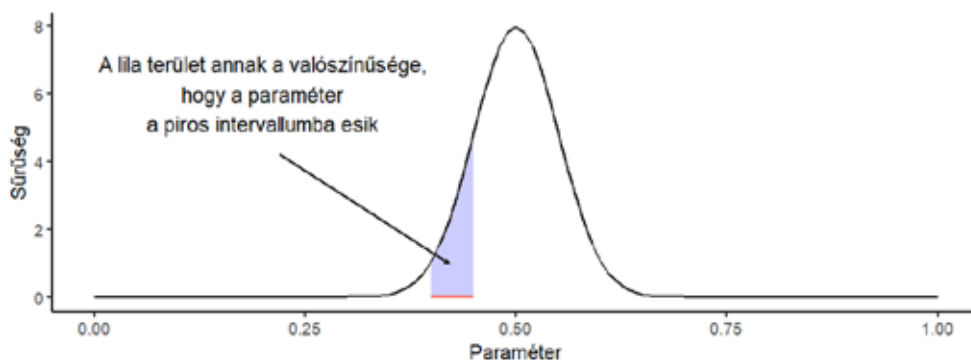
A Bayes-i modellezéshez három alapvető összetevőre van szükségünk: a modellre, az adatokra és az ún. *prior* ismeretekre. A prior információ sűríti magába a folyamattal kapcsolatos előzetes várakozásokat és ismeretanyagot. A prior három fő típusát különböztetjük meg. Ha nem áll rendelkezésünkre olyan információ, amelyet hasznos lenne beépíteni a modellbe, pl. egy új, eddig ismeretlen tudományterület feltérképezésekor, akkor ún. *nem informatív prior* választunk. Ez a prior nem hordoz információt előzetes várakozásokról, egyedül a rendelkezésre álló adatok alapján becsüljük meg a modell paramétereit. Abban az esetben is választhatunk nem informatív priort, ha

nagy mennyiségű adattal rendelkezünk, amelyből további feltételezések nélkül is hatékony becslés várható.

A második lehetőség az ún. *gyengén informatív prior* használata. Ennek szerepe csupán a paraméterek tartományának korlátozása, ill. a modell stabilizálása. Egy diagnosztikai eljárás érzékenységéről legtöbbször feltehető, hogy 50%-nál nem kisebb. Gyógyíthatatlan betegség modellezésekor kizárhatjuk azokat a paraméterértékeket, amelyek megengednék a gyógyulást. Az adatok heterogenitását jellemző varianciaparaméterek általában nem lehetnek nulla közelében és túl nagyok se, különben instabil és értékelhetetlen lehet az illesztett Bayes-i modell. A prior eloszlások harmadik típusa az *informatív prior*, mely lehetővé teszi, hogy megbízható előzetes ismereteket építsünk be a modellbe. Ilyenkor szakértők bevonásával, korábbi tanulmányok eredményei, tapasztalataink, ismereteink, várakozásaink alapján megadjuk, hogy a paraméterek nagy valószínűséggel milyen értéktartományba eshetnek. A paraméterek ún. *prior eloszlását* úgy szerkesztjük meg, hogy annak valószínűsége, hogy a paraméter egy adott intervallumba esik, a prior eloszlás adott intervallum fölé eső görbe alatti területével egyezzen meg. A függvénygörbe értékeit *sűrűségnek* nevezzük; nagy sűrűség a paraméterérték nagy valószínűségét jelzi (**2. ábra**).



2. ábra: Prior és poszterior eloszlás értelmezése

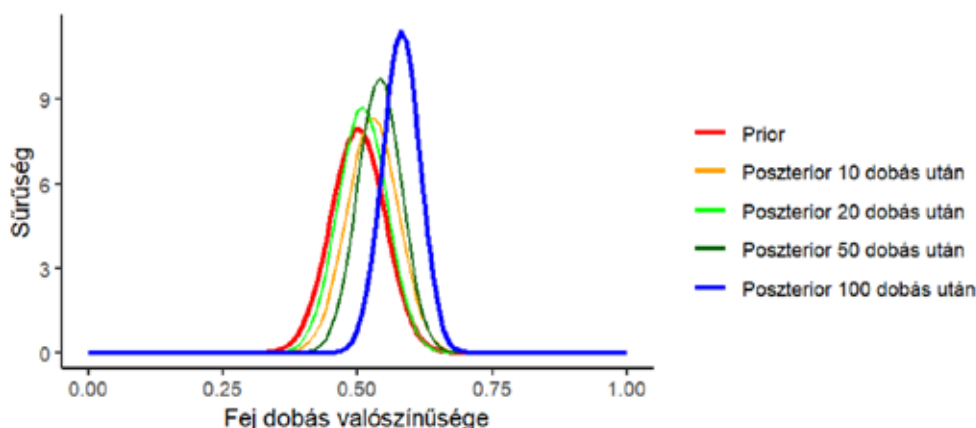


A prior eloszlást a Bayes-i modellben az adatok segítségével pontosítjuk, aktualizáljuk, így jutunk el az ún. *poszterior eloszláshoz*.

Érmefeldobás esetében, ha az az előfeltételezésünk, hogy a pénzérme szabályos, de persze előfordulhat-

nak benne apró gyártási hibák, használati kopások, válasszunk priornak egy olyan eloszlást (értéktartományt), mely szerint a fej-írás esélye kb. 50–50%, ettől el is térhet némiképp, de a nagyobb eltérés esélye kicsi (**3. ábra**).

3. ábra: Fej dobás valószínűségének Bayes-i becslése érmedobás során – szimuláció, cinkelt érme, fej valószínűsége: 0,6 (60%)



Prior feltételezés: szabályos érme

A keresett paramétert a prior eloszlásból kiindulva, az adatokban rejlő többletinformáció felhasználásával becsüljük meg. Becslésünk nem egy konkrét szám, hanem az előbbiekkal összhangban egy ún. *poszterior eloszlás* lesz. Annak a valószínűsége, hogy a valódi paraméter egy adott intervallumba esik, a poszterior eloszlás adott intervallum fölé eső görbe alatti területének kiszámításával adható meg (**2. ábra**).

Érmedobás esetében pl. azt mondja meg a poszterior eloszlás, hogy milyen eséllyel szabályos az érme, ill. milyen eséllyel tér el ettől egyik vagy másik irányban adott mértékkel.

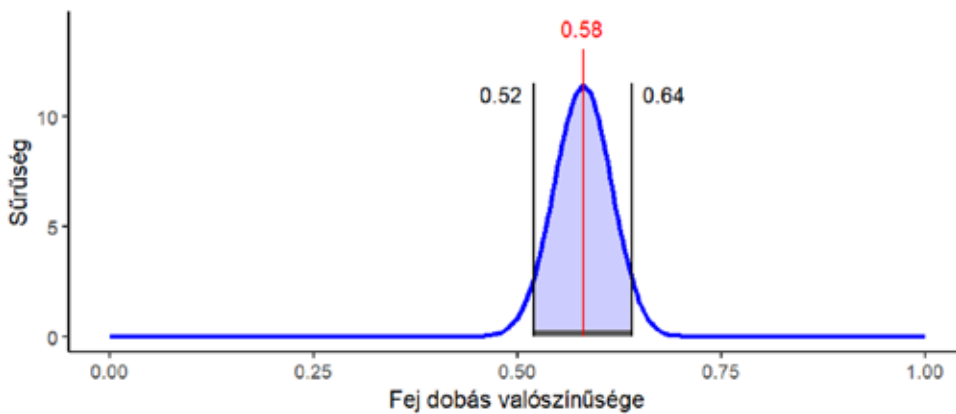
A **3. ábrán** láthatjuk, hogy az előfeltételezéseinket reprezentáló prior görbéből kiindulva a tapasztalatok felhalmozódásával (dobásszám növelése) egyre keskenyebb lesz a poszterior eloszlásunk, azaz nő a becslésünkbe vetett bizalom. Ezzel párhuzamosan a poszterior görbe a valódi paraméterérték felé tolódik el, azaz a valódi paraméter környékén található értékeknek lesz a legnagyobb a valószínűsége a



poszterior eloszlás alapján. A paraméterek poszterior eloszlását az adatok és a prior felhasználásával számítjuk ki. A könnyebb értelmezhetőség kedvéért a poszterior eloszlást egyetlen számba is sűrítjük, kiszámítva pl. az átlagos értékét vagy mediánját és az ehhez tartozó adott lefedettségű *kredibilis intervallumot* (K_{r1}). A *kredibilis intervallum lefedettsége* annak a valószínűsége, hogy a paraméter az adott intervallumba esik (**4. ábra**).



4. ábra: Poszterior eloszlás, becsült átlag és 95%-os kredibilis intervallum fej dobás valószínűségére 100 érmedobás után



Prior feltételezés: szabályos érme, valóság: cinkelt érme, fej valószínűsége 0.6 (60%), becsült átlag és kredibilis intervallum: 0.58 (95% Kr: 0.52-0.64)

A becslés eredményében megjelenik az adatokban rejlő többletinformáció, ismereteink gyarapodnak. A frissített poszterior ismerethalmaz felhasználható egy következő modell prior eloszlásaként. Ez a munkamódszer jól illeszkedik a tudományos megismerés természetes folyamatába. Előzetes ismereteinkből és az aktuális kutatásban tapasztaltakból születik meg az új elmélet, amelyet az újabb adatok tükrében ismét újra és újra felülvizsgálunk, ezzel finomítva lépésről-lépésre a vizsgált témakörrel alkotott képünket.

Az orvosi diagnosztika gondolkodásmódja is természetes módon Bayes-i szellemű. A diagnosztika fő kérdése, hogy adott tünetegyüttes ismeretében mi a valószínűsége az egyes kiváltó okoknak. Az esélyek latolgatásánál figyelembe vesszük a páciens jellemzőit, adottságait és a környezeti tényezőket is. Attól függően pl., hogy milyen idős, milyen korábbi betegségei voltak, továbbá milyen környezetből származik, más és más lehet egy-egy mögöttes ok valószínűsége. Pl. lázas, hasmenéses, bágyadt és levert, 6 hónapos kölyökutya esetében, amely egyébként rendszeres állategészségügyi ellátásban és az életkornak megfelelő ajánlott vakcinázásban

is részesült, első körben nem parvovírusos fertőzésre gyanakodunk. Azonban ugyanezekkel a tünetekkel egy tenyésztőtől most vásárolt, még semmilyen vakcinázásban nem részesülő, állatorvosnál először járó, 3 hónapos kutyakölyök esetében a parvovírus okozta megbetegedés valószínűségét jóval nagyobbra fogja értékelni a klinikus állatorvos.

Fertőző betegségek populációs részarányának, prevalenciájának becslésében a Bayes-i megközelítés használatát többnyire két fő érv motiválja. Egyrészt szeretnénk beépíteni a korábbi ismereteinket a becslés pontosításához. Ha ismerjük pl. a legutóbbi szűrésnél becsült prevalenciát, azt feltételezhetjük, hogy újra a korábbihoz hasonló értéket kapunk, de akár nagy változás is előfordulhat, kis eséllyel. Ezeket az ismereteket, várakozásokat építjük be priorként a modellbe. Másrészt a diagnosztikai tesztek jellemzőit, a szenzitivitást (érzékenységet) és a specifitást (fajlagosságot) is általában bizonytalanság terheli, hiszen azokat is kísérleti adatokból becsülik meg. Egy Bayes-i modell a diagnosztikai paraméterek bizonytalanságát is képes prior eloszlás formájában beépíteni.

