A bayesiánus kezdeti valószínűségek kimosódása frekventista objektív valószínűségre alapozva[1]

Bitai Tamás

ELTE BTK Filozófia Intézet Logika Tanszék

A tudományos hipotézisek bayesiánus konfirmációelmélete melletti egyik érv a kezdeti valószínűségek kimosódásának (*washing out of the priors*) érve, mely szerint egy objektív valószínűségi állítást kifejező H hipotézis szubjektív valószínűsége 1-hez konvergál, ha kísérleti eredményeink valóban H-nak megfelelőek, és a kísérletek hatására a szubjektív valószínűségeinket a bayesiánus frissítés szerint módosítjuk; mindez pedig attól függetlenül teljesül, hogy mi volt H kezdeti szubjektív valószínűsége (így a kezdeti valószínűség „kimosódik"). Az érv tehát a bayesiánus módszer azon előnyére hivatkozik, hogy ha egy valószínűségi hipotézis igaz, akkor a bayesiánus módszer szerint konfirmálódni is fog, függetlenül a bayesiánus ágens kezdeti hiteitől. Így, bár a bayesiánus konfirmáció szubjektív hitállapotokat alkalmaz, mégis rendelkezik a tudományos igazolástól elvárt objektivitással, hiszen az episztemikus ágensek szubjektív hiteinek különbségei a kísérletek hatására megszűnnek. A következőkben az érv részletes kifejtését John Earman és Wesley C. Salmon nyomán fogalmazzuk meg (Salmon et al. 1992, 2.8. szakasz).

Tekintsük a következő példát. Egy pénzérméről azt gyanítjuk, hogy mindkét felén fej van. Az érmét nem tudjuk megnézni, mert a szomszéd szobában van, viszont valaki sorozatban feldobja, és megmondja nekünk, hogy fej vagy írás lett az eredmény. Az egyszerűség kedvéért tegyük fel, hogy tudjuk, hogy ha az érmének nem mindkét felén fej van, akkor szabályos, azaz az egyik felén fej van, a másikon írás, és mindkettőt $\frac{1}{2}$ valószínűséggel lehet dobni, úgy, hogy az egyes dobások egymástól függetlenek. Tegyük fel továbbá, hogy mind a hipotézisünkben, mind a tagadásában valamilyen nem 0 mértékben hiszünk. Nézzük meg, hogyan változnak a szubjektív valószínűségeink, ha a hipotézisünk igaz, vagyis az érmével csak fejet lehet dobni. Hipotézisünket jelöljük H-val, ekkor $\neg H$ jelenti azt, hogy az érme szabályos; az egyes dobások lehetsé-

1 E. Szabó Lászlóval közös munka eredménye. Köszönöm Gömöri Márton javaslatait.

ges kimeneteleit pedig jelöljük $\langle r\rangle_n$-nel, $r \in \{\text{heads}, \text{tails}\}$, $n \in \{1, 2, \dots\}$. A bayesiánus konfirmációelmélet szerint a szubjektív valószínűségeink (hiteink mértékei) kezdetben egy $\mathbb{P}_0 = \mathbb{P}$ valószínűségi mértéket alkotnak, az n-edik kísérlet hatására pedig a kísérleti eredményre való kondicionálással frissítődnek:

$$\mathbb{P}_n(\cdot) = \mathbb{P}_{n-1}(\cdot \mid \langle r\rangle_n),$$

ahol

$$r = r(n) = \begin{cases} \text{heads}, & \text{ha az } n\text{-edik dobás fej,} \\ \text{tails}, & \text{ha az } n\text{-edik dobás írás.} \end{cases}$$

Felhasználva, hogy egymás utáni kondicionálások sorozata az állítások konjunkciójára való kondicionálással helyettesíthető:

$$\mathbb{P}_n(\cdot) = \mathbb{P}\left(\cdot \,\middle|\, \textstyle\bigwedge_{i=1}^n \langle r\rangle_i\right).$$

Ezt H-ra felírva, majd a Bayes-szabályt alkalmazva:

$$\mathbb{P}_n(H) = \frac{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, H\right)\mathbb{P}(H)}{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i\right)}.$$

A nevezőben a teljes valószínűség tételét alkalmazva a $\{H, \neg H\}$ egységfelbontásra:

$$\mathbb{P}_n(H) = \frac{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, H\right)\mathbb{P}(H)}{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, H\right)\mathbb{P}(H) + \mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, \neg H\right)\mathbb{P}(\neg H)}.$$

A számlálóban lévő kifejezéssel egyszerűsítve:

$$\mathbb{P}_n(H) = \frac{1}{1 + \dfrac{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, \neg H\right)}{\mathbb{P}\left(\bigwedge_{i=1}^n \langle r\rangle_i \,\middle|\, H\right)}\dfrac{\mathbb{P}(\neg H)}{\mathbb{P}(H)}}.$$

Ha H-nak megfelelően csupa fej dobás történik, akkor a fenti formulában $\langle r\rangle_i$ értéke minden i-re $\langle\text{heads}\rangle_i$ lesz. H szerint (csak fejet lehet dobni) egy n hosszú fej dobássorozat objektív valószínűsége 1, $\neg H$ szerint (az érme szabályos) $\left(\frac{1}{2}\right)^n$. Az objektív és szubjektív valószínűséget összekapcsoló David Lewis-féle *principal principle* szerint ekkor

$$\mathbb{P}(\textstyle\bigwedge_{i=1}^n \langle\text{heads}\rangle_i \mid H) = 1,$$

illetve

$$\mathbb{P}\left(\bigwedge_{i=1}^{n}\langle\text{heads}\rangle_i \mid \neg H\right) = \left(\tfrac{1}{2}\right)^n.$$

Ezeket beírva a fent $\mathbb{P}_n(H)$-ra levezetett formulába:

$$\mathbb{P}_n(H) = \frac{1}{1 + \frac{\left(\frac{1}{2}\right)^n}{1}\frac{\mathbb{P}(\neg H)}{\mathbb{P}(H)}},$$

ami 1-hez tart. Tehát valóban, ha a hipotézisnek megfelelő kísérletek történnek, akkor a hipotézis a kísérletsorozat határértékében konfirmálódik, függetlenül attól, hogy kezdetben milyen mértékben hittünk az érvényességében, feltéve, hogy ez a mérték nem 0 volt.

A fenti gondolatmenet hiányossága, hogy a konvergenciát csak egy példára vezeti le, azonban a teljes bizonyításhoz tetszőleges H és $\neg H$ esetén volna szükség a konvergencia megmutatására. A következőkben ezt fogjuk megtenni. Legyen H az a hipotézis, hogy a fej dobás objektív valószínűsége $p^{(H)}_{\text{heads}}$, az írásé $p^{(H)}_{\text{tails}}$; $\neg H$ pedig az a hipotézis, mely szerint ezek a valószínűségek $p^{(\neg H)}_{\text{heads}}$, illetve $p^{(\neg H)}_{\text{tails}}$. Míg Earman és Salmon példájában egyértelmű volt, mit jelent, hogy a kísérleti eredmények a H-nak megfelelőek, ebben az általános esetben nem nyilvánvaló ennek az értelmezése. Mi itt ezt az objektív valószínűség frekventista értelmezése szerint fogjuk megtenni. A frekventizmus tézise, hogy az objektív valószínűség a relatív gyakoriságok határértéke. Eszerint az, hogy egy kísérletsorozat H-nak megfelelő, azt jelenti, hogy benne a fejek, illetve írások relatív gyakorisága tart $p^{(H)}_{\text{heads}}$-hoz, illetve $p^{(H)}_{\text{tails}}$-hoz. $\mathbb{P}_n(H)$ 1-hez való konvergenciáját a példához hasonlóan most is a dobássorozatoknak a hipotézisekre vett feltételes valószínűségei hányadosának 0-hoz való konvergenciájára vezethetjük vissza. Jelöljük ezt a hányadost $D_n(H)$-val. A példában $D_n(H)$ értéke $\frac{\left(\frac{1}{2}\right)^n}{1}$ volt, általános esetben ez a következő lesz:

$$D_n(H) = \frac{\prod_{i=1}^{n} p^{(\neg H)}_{r(i)}}{\prod_{i=1}^{n} p^{(H)}_{r(i)}}, \text{ ahol } r(i) = \begin{cases} \text{heads}, & \text{ha az } i\text{-edik dobás fej,} \\ \text{tails}, & \text{ha az } i\text{-edik dobás írás.} \end{cases}$$

Bizonyításunkban a Kullback–Leibler-divergencia fogalmát fogjuk használni, amely H és $\neg H$ esetére a következőképpen van definiálva:

$$D_{\text{KL}}(H \parallel \neg H) = - \sum_{r \in \{\text{heads}, \text{tails}\}} p^{(H)}_r \log_2 \frac{p^{(\neg H)}_r}{p^{(H)}_r}.$$

A $D_n(H)$-ra felírt fentebbi összefüggésben logaritmusra áttérve:

$$\log_2 D_n(H) = \sum_{i=1}^{n} \log_2 \frac{p_{r(i)}^{(\neg H)}}{p_{r(i)}^{(H)}}.$$

Ennek az n tagú összegnek egy részében a fej dobás H, illetve $\neg H$ szerinti objektív valószínűségei fognak szerepelni, másik részében az íráséi; jelöljük ezen részek n-hez viszonyított arányának $p_{\text{heads}}^{(H)}$-tól, illetve $p_{\text{tails}}^{(H)}$-tól való eltérését d-vel:

$$\log_2 D_n(H) = n \underbrace{\left((p_{\text{heads}}^{(H)} + d) \log_2 \frac{p_{\text{heads}}^{(\neg H)}}{p_{\text{heads}}^{(H)}} + (p_{\text{tails}}^{(H)} - d) \log_2 \frac{p_{\text{tails}}^{(\neg H)}}{p_{\text{tails}}^{(H)}} \right)}_{D'_n(H)}.$$

Rögzítsünk egy tetszőleges $\varepsilon > 0$ számot. H frekventista értelmezése szerint a dobásokban a fejek, illetve írások relatív gyakorisága $p_{\text{heads}}^{(H)}$-hoz, illetve $p_{\text{tails}}^{(H)}$-hoz tart, így feltehetjük, hogy n már akkora, hogy

$$|d| < \frac{\varepsilon}{2 \max \left\{ \left| \log_2 \frac{p_{\text{heads}}^{(\neg H)}}{p_{\text{heads}}^{(H)}} \right|, \left| \log_2 \frac{p_{\text{tails}}^{(\neg H)}}{p_{\text{tails}}^{(H)}} \right| \right\}}.$$

Ekkor

$$\left| D'_n(H) - -D_{\text{KL}}(H \parallel \neg H) \right| < \varepsilon.$$

A Kullback–Leibler-divergencia a Gibbs-egyenlőtlenség szerint pozitív, így azt kaptuk, hogy $D'_n(H)$ egy negatív számhoz tart. Így $\log_2 D_n(H) \to -\infty$, következésképpen $D_n(H) \to 0$, amit bizonyítanunk kellett.

A fentiekben a kimosódás érvének egy teljes formális rekonstrukcióját adtuk meg. Az irodalomban szereplő formalizálásokról Earman ad tájékoztatást (Earman 1992, 6. fejezet, 3–5. szakaszok): ezek a formalizálások nem köteleződnek el az objektív valószínűség semmilyen értelmezése mellett, de csak majdnem biztos (azaz 1 valószínűségű, az objektív valószínűség értelmében) konvergenciát bizonyítanak (pl. a nagy számok erős törvényére hivatkoznak). Ezzel szemben a mi formalizálásunk az objektív valószínűség frekventista értelmezésével „biztos” (tényleges, feltétel nélküli) konvergenciát bizonyít. Ezt meggyőzőbb érvnek tartjuk a bayesiánus konfirmációelmélet mellett azok számára, akik az objektív valószínűség frekventista értelmezését elfogadják.

Irodalomjegyzék

SALMON, Merrilee H. et al. (1992): *Introduction to the Philosophy of Science.* Prentice-Hall.

EARMAN, John (1992): *Bayes or Bust? – A Critical Examination of Bayesian Confirmation Theory.* MIT Press.