INFORMÁCIÓELMÉLETI MÓDSZEREK A BIZTOSÍTÁSBAN[1]

KOMÁROMI ÉVA
Budapesti Közgazdaságtudományi Egyetem

Az információtartalom, entrópia, információdivergencia információelméleti fogalmak, de széles körű alkalmazással bírnak ezek a fogalmak a fizikában, biológiában, orvostudományban, mezőgazdaságban, regionális tervezésben, alakfelismerésben és természetesen a közgazdaságtan, pénzügy, biztosítás területén is. A szolgáltatások és díjak, tartalék, függőkár, megtartás stb. számításához elengedhetetlen a statisztikai módszerek alkalmazása, a számítógépek jelenléte egyben lehetővé és természetessé is teszi az aktuárius számára a rendszeres statisztikai kiértékelést. Ebben a dolgozatban a statisztika egyik olyan területével foglalkozunk, amelyben a feltételek melletti optimalizálás kiemelt jelentőséggel bír. Eleinte a használt fogalmakat tisztázzuk, a matematikai apparátust mutatjuk be, majd az alkalmazás lehetőségét három példán illusztráljuk.

1. Információtartalom

Tegyük fel, hogy ismerjük egy E esemény p bekövetkezési valószínűségét. Tegyük fel, hogy egy későbbi időpontban határozott és megbízható hírt kapunk arról, hogy E valóban bekövetkezett. Ha $p = 0,99$, ez a hír nem lep meg, vagyis a hír információtartalma igen kicsi, ha p értéke 1-hez közel van. Ha azonban 0-hoz közeli, például $p = 0,01$, akkor a hír meglepő lesz, hiszen gyakorlatilag bizonyos volt, hogy az esemény nem következik be, vagyis ekkor a hír információtartalma nagy. Az információtartalom tehát a valószínűség csökkenő függvénye. E természetes követelménynek, továbbá annak a követelménynek is, hogy független események esetében additív legyen, eleget tesz az *információtartalom* mérésére általában használt $h(p) = \ln\frac{1}{p} = -\ln p$ függvény. Megjegyezzük, hogy az információtartalom axiomatikus megalapozásával itt nem foglalkozunk, az érdeklődő olvasó ezt megtalálja például H. Theil könyvében [1].

Legyenek most egy kísérlet n lehetséges kimenetelének valószínűségei sorra $p_1, p_2, \ldots, p_n : p_i \geq 0, \quad i = 1, \ldots, n, \quad \sum p_i = 1$. Ha egy határozott és meg-

[1] Beérkezett 1996. október 3.

bízható hírt veszünk arról, hogy az i. esemény bekövetkezett, akkor ennek a hírnek az infomációtartalma $h(p_i) = -\ln p_i$. Mielőtt a hír megérkezett, nem tudhattuk, mekkora lesz az információtartalma, hiszen a $h(p_1), \ldots, h(p_n)$ számok bármelyike lehet. A hír megérkezése előtt is van azonban támpontunk: kiszámíthatjuk az érkező hír *várható információtartalmát*, azaz a

$$H(P) = H(p_1, \ldots, p_n) = \sum_{i=1}^{n} p_i h(p_i) = -\sum_{i=1}^{n} p_i \ln p_i$$

értéket. A várható információtartalom egyúttal a szóban forgó valószínűség-eloszlásban bennefoglalt bizonytalanság, rendezetlenség egyik alkalmazható mértéke is. A $H(P)$ függvény a bizonytalanság *Shannon* által bevezetett mértéke.

2. A valószínűségeloszlásban megjelenő bizonytalanság mértéke: az entrópia

Ésszerű feltételezés, hogy egy kísérlet lehetséges kimeneteleit illető bizonytalanság H mértéke eleget tegyen a következő kívánalmaknak:

(i) H a lehetséges kimenetelek valószínűségeinek függvénye:

$$H = H_n(P) = H_n(p_1, \ldots, p_n) \, .$$

(ii) A $p_1, p_2, \ldots, p_n$ valószínűségeknek folytonos függvénye: ezek kis változásai H_n kis változását idézik elő.

(iii) Nem változik, ha a lehetetlen kimenetelt hozzávesszük a valószínűségi sémához:

$$H_{n+1}(p_1, \ldots, p_n, 0) = H_n(p_1, \ldots, p_n) \, .$$

(iv) Nem változik, ha a lehetséges kimenetelek sorrendjét megváltoztatjuk, azaz H_n az argumentumainak szimmetrikus függvénye.

(v) Minimális (zéró) értékű, ha a kimenetelt illetően nincs bizonytalanság:

$$H_n(p_1, \ldots, p_n) = 0, \quad \text{ha} \quad p_i = 1 \qquad (i = 1, \ldots, n) \, .$$

(vi) Maximális értékű, ha az összes kimenetel egyenlően valószínű, azaz a valószínűségek nem adnak támpontot, teljes a bizonytalanság azt illetően, hogy milyen esemény következik be: H_n felveszi a maximumát, ha

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n} \, .$$

(vii) H_n maximális értéke n értékével növekszik.

(viii) Két független valószínűségi eloszlás esetén, ha $A_1, \ldots, A_n$ illetve $B_1, \ldots, B_m$ jelölnek lehetséges kimeneteleket és $p_1, \ldots, p_n$ illetve $q_1, \ldots, q_m$ a kapcsolódó valószínűségeket, akkor az A_iB_j lehetséges kimenetelekkel és p_iq_j valószínűségekkel bíró kísérlet ($i = 1, \ldots, n$; $j = 1, \ldots, m$) bizonytalanságát mérő $H_{n+m}(P \cup Q)$ függvényre fennáll, hogy

$$H_{nm}(P \cup Q) = H_n(P) + H_m(Q).$$

Egy valószínűségeloszlás bizonytalanságának a mértékét *az eloszlás entrópiájának* nevezzük. Itt az "entrópia" kifejezés információelméleti és nem termodinamikai eredetű.

A bizonytalanság Shannon által bevezetett (várható információtartalom) függvénye kielégíti a felsorolt nyolc tulajdonságot: folytonos és szimmetrikus függvény, és ha a $0 \ln 0 = 0$ konvenciót alkalmazzuk, akkor teljesíti a (iii) tulajdonságot. Mivel $-\sum_{i=1}^{n} p_i \ln p_i$ konkáv függvény, ezért lokális maximumhelye: $p_i = \frac{1}{n}$, $i = 1, \ldots, n$ egyúttal globális is, és maximális értéke $\ln n$. Maximális értéke n növekedésével nő. Ha a valószínűségek egyike 1, akkor $H_n = 0$, ez tehát a nemnegatív H függvény minimuma. Végül, független valószínűségi változók esetén

$$\begin{aligned}
H_{nm}(P \cup Q) &= -\sum_{i=1}^{n}\sum_{j=1}^{m}(p_iq_j)\ln(p_iq_j) \\
&= -\sum_{i=1}^{n} p_i \left[\sum_{j=1}^{m} q_j \ln q_j\right] - \sum_{j=1}^{m} q_j \left[\sum_{i=1}^{n} p_i \ln p_i\right] \\
&= \sum_{i=1}^{n} p_i H_m(Q) + \sum_{j=1}^{m} q_j H_n(P).
\end{aligned}$$

Az információelméletben nem a felsorolt nyolc követelmény alkotja az egyetlen axiómarendszert a bizonytalanság mértékét illetően, ennek megfelelően nem a Shannon-féle entrópia az egyetlen entrópia-függvény. Más axiómarendszernek más függvények tesznek eleget, erre vonatkozó utalásokat ld. pl. Mathai és Rathie [3] illetve Kapur [2] könyvében. A különböző típusú entrópiák közötti összefüggést és az entrópiafogalom hírközlésben való alkalmazását F. M. Reza [4] részletesen tárgyalja.

3. Információ-divergencia

Tegyük fel ismét, hogy van n eseményünk, amelyek teljes eseményrendszert alkotnak és valószínűségeik rendre $\alpha_1, \alpha_2, \ldots, \alpha_n$. Most és a továbbiakban feltesszük, hogy $\alpha_j > 0$, $j = 1, \ldots, n$. Az $\alpha_1, \alpha_2, \ldots, \alpha_n$ értékeket előzetes vagy *a priori* valószínűségeknek nevezzük. Beérkezik a hír, de nem arról, hogy valamelyik esemény bekövetkezett volna, hanem arról, hogy az esélyek megváltoztak: némelyik esemény valószínűbbé vált, más meg valószínűtlenebbé. A hír az előzetes valószínűségeket a $p_1, p_2, \ldots, p_n$ utólagos vagy *a posteriori* valószínűségekké változtatja. Ha egyik p_i értéke 1, a többi 0, akkor a hírben közvetített információ-tartalom $-\ln \alpha_i$. Ha a hír csak az i. eseményre szorítkozik, akkor a hírben közvetített információ az előzetes és az utólagos valószínűségekben foglalt információ-tartalmak különbsége: $\ln \frac{p_i}{\alpha_i}$. A hír azonban nem szorítkozhat egyetlen eseményre, ellenkezőleg, azt állítja, hogy minden eseménynek megvan a saját utólagos valószínűsége. Így a hírben közvetített várható információtartalom: $I_n(P) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{\alpha_i}$, amelyet *információ divergenciának* vagy *I-divergenciá*nak neveznek. Jelölésére az $I(\alpha; P)$ is használatos. Felhívjuk az olvasó figyelmét Csiszár dolgozatára [5] az információ divergencia és az optimalizációs problémák közötti kapcsolatról.

Az $I_n(P)$ várható információtartalom nemnegatív. Ez az állítás a számtani és mértani közép között fennálló egyenlőtlenségből adódik a következőképpen:

$$1 = \sum_{i=1}^{n} p_i \frac{\alpha_i}{p_i} \geq \prod_{i=1}^{n} \left(\frac{\alpha_i}{p_i} \right)^{p_i}$$

$$0 \geq \sum_{i=1}^{n} p_i \ln \frac{\alpha_i}{p_i}, \qquad \text{vagyis} \qquad 0 \leq \sum_{i=1}^{n} p_i \ln \frac{p_i}{\alpha_i}$$

és egyenlőség áll fenn akkor és csak akkor, ha $p_i = \alpha_i$, $i = 1, \ldots, n$. Megjegyzendő, hogy $I_n(P)$ értéke végtelen nagy is lehet akkor, ha valamely hírre $p_i > 0$ *és* $\alpha_i = 0$. Megjegyzendő továbbá, hogy teljes előzetes bizonytalanság esetén, amikor tehát $\alpha_i = \frac{1}{n}$ minden i indexre, akkor a várható információ: $I_n(P) = \ln n - \sum_{i=1}^{n} p_i \ln \frac{1}{p_i} = \ln n - H_n(P)$.

4. A maximum-entrópia elv

A várható információ most bevezetett fogalma és jelentése alapján kibővíthetjük az entrópia fogalmát is. A Shannon-féle entrópia maximális értékű, ha valamennyi lehetséges kimenetel egyenlően valószínű. Ha azonban okunk van azt feltételezni, hogy adott egy *a priori* $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ eloszlás ($\alpha_i \geq 0$, $\sum \alpha_i = 1$), akkor egy másik entrópiát vezethetünk be a következő összefüggés formájában:

$$B(P) = -\left[\sum_{i=1}^{n} p_i \ln \frac{p_i}{\alpha_i}\right] - \ln\left(\min\{\alpha_1, \ldots, \alpha_n\}\right)$$

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \alpha_i = 1, \quad p_i \geq 0, \alpha_i \geq 0,$$

amely az $I_n(P) = \sum\limits_{i=1}^{n} p_i \ln \frac{p_i}{\alpha_i}$ helyettesítéssel

$$B(P) = -I_n(P) - \ln\left(\min\{\alpha_1, \ldots, \alpha_n\}\right)$$

alakban írható. Kapur említett könyvében $B(P)$-t *Bayes-féle entrópiának* nevezi, mivel a kifejezés az *a priori* és *a posteriori* valószínűségeloszlások egy függvénye. Értéke maximális, ha az $I_n(P)$ információ-divergencia minimális és minimális, ha $I_n(P)$ maximális.

Ha csak az előzetes $(\alpha_1, \ldots, \alpha_n)$ valószínűségeloszlást ismerjük, a $P = (p_1, p_2, \ldots, p_n)$ valószínűségeloszlásról azt kell feltételeznünk, hogy megegyezik az előzetes valószínűségeloszlással, azaz minimalizálja az információ-divergenciát. Ha azonban más információ is rendelkezésünkre áll, például a megjósolandó eloszlás várható értéke vagy szórása, akkor nem választhatjuk az előzetes eloszlást, ha az nem elégíti ki a szóban forgó feltételt. Ekkor olyan P utólagos valószínűségeloszlást kell választanunk, amely az adott feltétel(ek) mellett a "legközelebb áll" az előzetes valószínűségeloszláshoz: minimalizálja az információ-divergenciát, vagyis maximalizálja a Bayes entrópiát. Ez a *"maximum-entrópia elv"*. Azt a valószínűségeloszlást, amelyet a maximum-entrópia elv alkalmazásával számítunk, maximum-entrópia eloszlás, legvalószínűbb eloszlás, legegyenletesebb eloszlás és egyéb neveken is említik.

A Bayes-entrópia a megszámlálhatóan végtelen lehetséges kimenetelt tartalmazó kísérletre a

$$B = -\sum_{i=1}^{\infty} p_i \ln \frac{p_i}{\alpha_i}$$

formulával adható meg. Hasonlóan, folytonos esetben

$$B = -\int\limits_{\mathrm{supp}\,\alpha}^{\infty} f(x) \ln \frac{f(x)}{\alpha(x)} dx,$$

ahol $f(x)$ és $\alpha(x)$ az utólagos (*a posteriori*) és előzetes (*a priori*) valószínűségi eloszlás sűrűségfüggvénye. Mindkét esetben a minimalizálandó I információ-divergencia a Bayes-entrópia negatívja, a P valószínűségeloszlás ill. $f(x)$ sűrűségfüggvény konvex függvénye. Az információ-divergencia minimalizálása rendszerint loglineáris model, amelynek megoldása konvex programozási eszközökkel történhet.

5. Az I-divergencia minimalizálása, mint konvex programozási feladat

A minimalizációs feladat, amellyel a konvex programozás foglalkozik, a következő:

$$\begin{aligned} \theta(x) &\rightarrow \min \\ g_i(x) &\leq 0, \quad i = 1, \dots, m; \quad x \in R^n \end{aligned}$$

E feladathoz kapcsolódik a

$$\begin{aligned} \nabla\theta(x) + \sum_{i=1}^{n} u_i \nabla g_i(x) &= 0, \quad u_i \geq 0, \qquad i = 1, \dots, m; \\ \sum_{i=1}^{n} u_i g_i(x) &= 0, \quad g_i(x) \leq 0, \quad i = 1, \dots, m; \end{aligned}$$

Kuhn-Tucker feladat, amelynek a megfogalmazásában feltételezzük, hogy θ és g_i differenciálható függvények. Az u_i változókat Lagrange-szorzóknak nevezzük, a $\theta(x) + \sum u_i g_i(x)$ függvényt pedig Lagrange függvénynek. A Kuhn-Tucker feladat első n feltétele azt írja elő, hogy a Lagrange függvény parciális deriváltjai legyenek 0 értékűek. A következő m feltétel a Lagrange szorzók nemnegativitását illeti, az utolsó m feltétel a minimalizációs feladat feltételei. A közbülső feltétel az egyensúlyi feltétel: mivel $u_i \geq 0$ és $g_i \leq 0, \quad i = 1, \dots, m$, ezért szorzataik összege nempozitív, és 0 csak akkor, ha az összeg minden tagja 0, azaz, $u_i = 0$ ha $g_i(x) < 0$. Ha egyenlőtlenségi feltételek helyett egyenlőségi feltételek szerepelnek a minimalizációs feladatban, akkor a megfelelő u_i Lagrange szorzók előjelkötetlenek és a megfelelő tagok az egyensúlyi feltételben 0 értékűek. A Kuhn-Tucker feladathoz és az alább bemutatott nyeregpont feladathoz kapcsolódó tételeket, eredményeket ld. pl. Mangasarian [6] könyvében.

Kuhn-Tucker optimalitási tétel: Tegyük fel, hogy a θ és g_i függvények differenciálhatók az $x^\circ \in R^n$ pontban. (1) Tegyük fel, hogy θ és g_i konvexek. Ha (x°, u°) megoldása a Kuhn-Tucker feladatnak, akkor x° optimális megoldása a minimalizációs feladatnak. (2) Tegyük fel, hogy x° optimális megoldása a minimalizációs feladatnak. Tegyük fel, hogy a feladat feltételrendszere kielégíti valamelyik regularitási feltételt. Akkor létezik olyan $u^\circ \in R^m$, hogy (x°, u°) kielégíti a Kuhn-Tucker feladatot.

Ezt a tételt abban az esetben fogjuk alkalmazni, amikor a szóban forgó valószínűség-eloszlás diszkrét. Az I-divergencia konvex függvény. Minimalizálásakor azon feltételek mindig megjelennek, amelyek kikötik, hogy a p_j változók nemnegatívok és összegük 1. A többi feltételek rendszerint valamely momentum értékére, pl. a várható értékre vonatkoznak egyenlőségi feltételek formájában. A tétel egyenlőségi feltételek esetén akkor alkalmazható, ha mind a g_i, mind a $-g_i$ függvények konvexek, azaz g_i a p_j változók lineáris függvénye, amely a felsorolt esetekben fennáll. A regularitási feltételek közül a leggyakrabban alkalmazottak a Slater feltétel, az Arrow-Hurwitz-Uzawa feltétel, a Kuhn-Tucker feltétel és a linearitás. Az általunk itt tárgyalt esetekben a feltételek lineárisak, így a tétel második állítása is alkalmazható.

A minimalizációs problémához kapcsolódik az alábbi, *Kuhn-Tucker nyeregpont feladat* is, amelyben a nyeregfüggvény az említett Lagrange függvény:

Keressünk olyan $x^\circ = (x_1^\circ, \ldots, x_n^\circ)$ és $u^\circ = (u_1^\circ, \ldots, u_m^\circ) \geq 0$ vektorokat, amelyekre fennáll, hogy

$$\theta(x^\circ) + \sum_{i=1}^{m} u_i g_i(x^\circ) \leq \theta(x^\circ) + \sum_{i=1}^{m} u_i^\circ g_i(x^\circ) \leq \theta(x) + \sum_{i=1}^{m} u_i^\circ g_i(x)$$

minden $u \geq 0$, $u \in R^m$, $x \in R^n$ mellett.

Kuhn-Tucker nyeregpont optimalitási tétel: (1) Ha (x°, u°) megoldása a Kuhn-Tucker nyeregpont feladatnak, akkor x° optimális megoldása a minimalizációs feladatnak. (2) Ha x° optimális megoldása a minimalizációs feladatnak és $\exists \widetilde{x} \in R^n$, amelyre $g_i(\widetilde{x}) > 0$ ha a g_i függvény nemlineáris, akkor $\exists u^\circ$, hogy (x°, u°) megoldása a Kuhn-Tucker nyeregpont feladatnak.

Ezt a tételt abban az esetben fogjuk alkalmazni, amikor a szóban forgó valószínűség-eloszlás folytonos. Vizsgáljuk először a diszkrét esetet. Feladatunk tehát

$$\begin{aligned} I_n(P) &= \sum_{j=1}^{n} p_j \ln \frac{p_j}{\alpha_j} \to \min \\ \sum_{j=1}^{n} a_{ij} p_j &= b_i, \quad i = 1, \ldots, m; \\ P &= (p_1, \ldots, p_n) \geq 0 \end{aligned}$$

alakú, ahol az A $m \times n$ méretű mátrix és a b m-méretű vektor adottak. Mivel a feltételek között a $\sum p_j = 1$ szükségképpen megjelenik, ezért az optimális megoldás (ha létezik) nem változik, ha a célfüggvényből levonjuk $\sum p_j$ -t, amit célszerűségi okokból megteszünk. Célfüggvényünk tehát:

$$\sum_{j=1}^{n} p_j \ln \frac{p_j}{\alpha_j} - \sum_{j=1}^{n} p_j \to \min .$$

Ehhez a feladathoz kapcsolódó Kuhn-Tucker feladat a következő: Keresünk olyan $P^\circ = (p_1^\circ, \ldots, p_n^\circ)$ és $u^\circ = (u_1^\circ, \ldots, u_m^\circ)$ értékeket, amelyek megoldják az alábbi feladatot:

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} p_j &= b_i, \quad i = 1, \ldots, m; \\ (p_1, \ldots, p_n) &\geq 0; \\ \ln p_j - \ln \alpha_j + \sum_{i=1}^{m} u_i a_{ij} &\geq 0, \quad j = 1, \ldots, n; \\ \sum_{j=1}^{n} p_j \left(\ln p_j - \ln \alpha_j + \sum_{i=1}^{m} u_i a_{ij} \right) &= 0 \end{aligned}$$

A Kuhn-Tucker optimalitási tétel szerint minimalizációs feladatunknak pontosan akkor van optimális megoldása, ha a leírt kapcsolódó Kuhn-Tucker feladatnak van (P°, u°) megoldása. Alkalmazva a $z_i = -u_i, i = 1, \ldots, m$ helyettesítést, a feladat P megoldását a következő ún. multiplikatív alakban:

$$p_j = \alpha_j e^{\sum z_i a_{ij}} = \alpha_j \prod_{i=1}^{m} \{e^{a_{ij}}\}^{z_i} ,$$

illetve a vele ekvivalens

$$\ln \frac{p_j}{\alpha_j} = \sum_{i=1}^{m} z_i a_{ij}$$

loglineáris alakban keressük, amely a fenti egyensúlyi feltétellel együtt hangsúlyozza az utólagos valószínűség azon tulajdonságát is, hogy valamelyik eleme a valószínűség-eloszlásnak pozitív értékű pontosan akkor, amikor az előzetes valószínűség megfelelő eleme pozitív értékű.

Írjuk most fel a minimalizációs feladatunk *duálisát*, és nézzük meg, mit mond a dualitási tétel a minimalizációs feladat optimális megoldásáról:

$$\sum_{j=1}^{n} p_j \ln \frac{p_j}{\alpha_j} - \sum_{j=1}^{n} p_j - \sum_{i=1}^{m} z_i \Big(\sum_{j=1}^{n} a_{ij} p_j - b_i \Big) - \sum_{j=1}^{n} p_j \left(\ln \frac{p_j}{\alpha_j} - \sum_{i=1}^{m} z_i a_{ij} \right) \to \max$$

$$\ln \frac{p_j}{\alpha_j} - \sum_{i=1}^{m} z_i a_{ij} \geq 0, \qquad j = 1, \ldots, n.$$

A feltételek így is felírhatók

$$p_j \geq \alpha_j e^{\sum z_i a_{ij}}, \quad j = 1, \ldots, n.$$

Vegyük észre, hogy a maximalizálandó célfüggvény összevonások után a következő:

$$-\sum_{j=1}^{n} p_j + \sum_{i=1}^{m} z_i b_i.$$

Vegyük továbbá észre, hogy a feladat optimális megoldására a feltételek egyenlőséggel kell, hogy teljesüljenek. Alakítsuk át a duális feladatot minimalizációs feladattá ez utóbbi észrevétel figyelembevételével. Ekkor az alábbi feltétel nélküli optimalizációs feladatot kapjuk:

$$\sum_{j=1}^{n} \alpha_j e^{\sum_{i=1}^{m} z_i a_{ij}} - bz \to \min$$

amely csak a $(z_1, \ldots, z_m)$ duális változók függvénye. A gyenge dualitási tétel értelmében a primál és duál feladat célfüggvényértékei között fennáll a

$$\sum_{j=1}^{n} \alpha_j e^{\sum_{i=1}^{m} z_i a_{ij}} - bz \geq -\sum_{j=1}^{n} p_j \ln \frac{p_j}{\alpha_j} + \sum_{j=1}^{n} p_j$$

egyenlőtlenség, amely egyenlőséggel teljesül, ha $p_j = \alpha_j e^{\sum_{i=1}^{m} z_i a_{ij}}$, $j = 1, \ldots, n$. A dualitási tétel szerint tehát, ha a primál feladatnak van optimális megoldása, akkor az ezen multiplikatív illetve az ekvivalens loglineáris formában írható fel.

Tekintsük most a folytonos esetet. Feladatunk ekkor

$$\begin{aligned}
\int_{\operatorname{supp} \alpha} f(x) \ln \frac{f(x)}{\alpha(x)} dx &\to \min \\
\int_{\operatorname{supp} \alpha} f(x) dx &= b_1 = 1, \\
f(x) &\geq 0, \\
\int_{\operatorname{supp} \alpha} a_i(x) f(x) dx &= b_i, \quad i = 2, \ldots, m,.
\end{aligned}$$

ahol supp α az $\alpha(x)$ sűrűségfüggvény tartója (az R azon legszűkebb részhalmaza, amelynek valószínűségi mértéke 1). E feladat Lagrange függvénye:

$$L = \int\limits_{\operatorname{supp}\alpha} f(x)\ln\frac{f(x)}{\alpha(x)}dx + \sum_{i=1}^{m} u_i \left(\int\limits_{\operatorname{supp}\alpha} a_i(x)f(x)dx - b_i \right)$$

ahol alkalmaztuk az $a_1(x) = 1$ jelölést. Rögzített u_i értékek esetén az L függvény az $f(x)$ függvényben minimalizálandó, a minimalizálás szempontjából pedig a $\sum\limits_{i=1}^{m} u_i b_i$, konstans lévén, elhagyható. Így a minimalizálandó függvény:

$$\int\limits_{-\infty}^{\infty} f(x)\left(\ln\frac{f(x)}{\alpha(x)} + \sum_{i=1}^{m} u_i a_i(x)\right)dx$$

$$= \int\limits_{-\infty}^{\infty} f(x)\left\{\ln\left[\frac{f(x)}{\alpha(x)}e^{\sum\limits_{i=1}^{m} u_i a_i(x)}\right]\right\}dx$$

$$\geq \int\limits_{-\infty}^{\infty} f(x)\left\{1 - \frac{\alpha(x)}{f(x)}e^{-\sum\limits_{i=1}^{m} u_i a_i(x)}\right\}dx$$

mivel $\ln Z \leq Z - 1$ és egyenlőség pontosan akkor áll fenn, ha $Z = 1$, vagyis amikor

$$f(x) = \alpha(x)e^{-\sum\limits_{i=1}^{m} u_i a_i(x)}$$

alakú, azaz, $z_i = -u_i$ helyettesítéssel, amikor

$$f(x) = \alpha(x)\prod_{i=1}^{m}\left(e^{a_i(x)}\right)^{z_i},$$

másként:

$$\ln\frac{f(x)}{\alpha(x)} = \sum_{i=1}^{m} z_i a_i(x)$$

alakú, így az eredményül kapott reprezentáció $f(x)$ számára ismét a multiplikatív, illetve a loglineáris forma.

Megjegyzendő, hogy ha előzetes valószínűségeloszlást nem veszünk számításba, akkor a fenti összefüggésekben az $\alpha_j = 1$, $j = 1, \ldots, n$ illetve $\alpha(x) \equiv 1$ helyettesítés alkalmazandó.

6. Entrópia-maximalizálásra épülő biztosításmatematikai modellek

Standard vagy referenciaeloszlásokat, statisztikákat kis biztosítótársaságok, vagy nagy biztosítótársaságok speciális módozataikban, nem mindig alkalmazhatnak rutinszerűen, mert klientúrájuk összetétele nem feltétlenül tükrözi annak a népességnek az összetételét, amelyre a statisztika készült. A probléma ekkor az, hogyan lehet elméletileg megalapozottan összehasonlítani és igazítani a referenciaeloszlást úgy, hogy a létrehozott eloszlás megfeleljen a szóban forgó biztosítótársaság klientúrájának, de a referenciaeloszláshoz a lehető legközelebb legyen. Az alábbi két problémával Brockett [7] illetve Brockett és Cox [8] foglalkoztak, a közölt megoldások tőlük származnak.

Munkaképtelenség időtartamának az eloszlása

Az egyszerűség kedvéért tegyük fel, hogy egyetlen feltétel van: a kliens várható munkaképtelenségének az időtartama: $\mu = 21$ nap, szemben a referenciatáblában feltüntetett várható időtartammal. A referenciatábla a $q_1, \ldots, q_n$ értékekből áll, ahol q_j annak a valószínűsége, hogy a munkaképtelenség j napig tart. A feladat olyan, a munkaképtelenség időtartamára vonatkozó táblát létrehozni, amely a referenciatáblától a lehető legkisebb mértékben különbözik, de teljesíti a $\mu = 21$ nap feltételt. A probléma megoldása többek között a díjszabás megállapításához is szükséges.

Jelölje q_i annak a valószínűségét, hogy egy ügyfél munkaképtelensége i napon át tart, a referenciatábla szerint. Jelölje p_i az ismeretlen valószínűségét annak, hogy az ügyfél munkaképtelenségének időtartama i nap lesz. A $\{p_i\}$ valószínűségeket az alábbi feladat megoldásaként kapjuk:

$$\sum_{i=1}^{\omega} p_i \ln \frac{p_i}{q_i} \quad \longrightarrow \quad \min$$

$$\sum_{i=1}^{\omega} p_i \quad = \quad 1, \quad \sum_{i=1}^{\omega} i p_i = 21, \quad p_i \geq 0, \quad i = 1, \ldots, \omega.$$

A fentiek szerint ennek a problémának a megoldása a következő alakú:

$$p_i = q_i e^{z_1 + i z_2} = q_i (e^{z_1})(e^{z_2})^i$$

Mivel a célfüggvény szigorúan konvex, ezért a minimalizációs feladatnak egyetlen optimális megoldása van, és így a p_i-re kapott kifejezés behelyettesítése után a feltételeket képviselő két egyenletből álló kétismeretlenes egyenletrendszernek egyetlen (z_1, z_2) megoldása van, amelyet iterációs eljárással vagy más módszerrel kiszámíthatunk. Brockett számítási eredményeit és az eredményül kapott valószínűség-eloszlást a hivatkozott dolgozata táblázatos formában bemutatja.

A halandósági táblák kiigazítása

Gyakori eset, hogy egy speciális ügyfélre vonatkozóan az aktuáriusnak el kell térnie számításaiban a standard halandósági táblától. Példaként tekintsünk egy 50 éves, bizonytalan egészségi állapotú ügyfelet, aki életbiztosítást is magában foglaló biztosítást kíván kötni. Az aktuárius szükségképpen választ egy standard halandósági táblát, amelyet azonban módosítania kell, ha az orvos az ügyfél hátralévő μ várható életkorát a halandósági táblában szereplő e_{50} értéktől különbözőnek ítéli. Tegyük fel, hogy az orvos szerint $\mu = 9 \neq e_{50}$. Tegyük fel továbbá, hogy emellett az orvos 50% esélyt ad annak, hogy az illető az 55. és 65. születésnapja között hal meg. Hogyan alakítsa az aktuárius a rendelkezésére álló halandósági táblát, hogy az a lehető legkevésbé térjen el a választott standard halandósági táblától, de mégis illeszkedjen az ügyfélnek az orvosi véleményben megfogalmazott halandósági tulajdonságaihoz – ezt fogalmazzuk meg az alábbi modellben, amelyben q_i jelöli a standard halálozási valószínűségeket, p_i pedig a számítandó halálozási valószínűségeket:

$$\begin{aligned}
\sum_{i=0}^{\omega} p_i \ln \frac{p_i}{q_i} \quad &\longrightarrow \quad \min \\
\sum_{i=0}^{\omega} p_i \quad &= \quad 1, \qquad \sum_{i=0}^{\omega} i p_i = 9, \\
\sum_{i=5}^{14} p_i \quad &= \quad 0,5, \qquad p_i \geq 0, \qquad i = 0, 1, \ldots, \omega\,.
\end{aligned}$$

A feladat feltételeit kielégítő p_i értékeket a

$$p_i = q_i e^{z_1 + i z_2 + a(i) z_3}$$

formában keressük, ahol

$$a(i) = \begin{cases} 0, & \text{ha } i < 5; \\ 1, & \text{ha } 5 \leq i \leq 14 \\ 0, & \text{ha } i > 14 \end{cases}$$

A feladat megoldása ismét egy háromismeretlenes, 3 egyenletet tartalmazó egyenletrendszer megoldását jelenti. Az eredmény Brockett és Cox [8] dolgozatában az ott közölt halandósági tábla alapján a következő:

$$p_i = \begin{cases} q_i(1,342125)(0,96353)^i, & \text{ha } i < 5 \\ q_i(1,891812)(0,96353)^i, & \text{ha } 5 \leq i \leq 14 \\ q_i(1,342125)(0,96353)^i, & \text{ha } i > 14 \end{cases}$$

Egyenlőtlenségi feltételeket is tartalmazó feladatok találhatók még Charnes et al. [9] dolgozatában.

A kárnagyság eloszlásának számítása információelméleti módszerrel

Legyen f(t) a sűrűségfüggvénye az egy bizonyos a érték fölötti kárigény nagyságának. Akkor

$$\int_a^\infty f(t)dt = 1.$$

A szimmetrikustól nagyon eltérő eloszlásnak esetleg nincs véges várható értéke, ezért az

$$\int_a^\infty \ln t f(t)dt = \ln b$$

feltételt támasztjuk. Ekkor az entrópiafüggvény maximalizálása két feltétel mellett történik, az $f(t)$ sűrűségfüggvény alakja a következő:

$$f(t) = e^{z_1 + z_2 \ln t}.$$

Használjuk fel az első feltételt:

$$\begin{aligned} 1 &= \int_a^\infty e^{z_1+z_2 \ln t} dt = e^{z_1} \int_a^\infty t^{z_2} dt \\ &= e^{z_1} \frac{a^{z_2+1}}{-z_2 - 1}, \quad \text{ahol} \quad z_2 < -1\,. \end{aligned}$$

Alkalmazva az $\alpha = -z_2$ helyettesítést, azt kapjuk, hogy $e^{z_1} = (\alpha - 1)e^{\alpha-1}$. Az eljárás tehát az

$$f(t) = (\alpha - 1)a^{\alpha-1}t^{-\alpha}, t \geq a, \alpha > 1$$

sűrűségfüggvényhez vezet, ami a *Pareto* eloszlás sűrűségfüggvénye. Ha $\alpha < 2$, akkor a kárnagyság várható értéke nem véges, de a $\ln t$ várható értéke véges. Ha $\alpha > 2$, akkor mindkét várható érték véges.

Irodalom

1. Theil, H.: Közgazdaságtan és információelmélet, Közgazdasági és Jogi Könyvkiadó, Budapest (1970).
2. Kapur, J. N.: Maximum-Entropy Models in Science and Engineering, John Wiley & Sons, Inc., New York (1989).

3. Mathai A. M., Rathie P. N.: Basic Concepts in Information Theory and Statistics, Wiley Eastern Limited, New Delhi (1975).
4. Reza, F. M.: Bevezetés az információelméletbe, Műszaki Könyvkiadó, Budapest (1966).
5. Csiszár, I.: I-Divergence Geometry of Probability Distribution and Minimization Problems, Annals of Probability 3 (1975), pp. 146-158.
6. Mangasarian, O. L.: Nonlinear Programming, McGraw-Hill Book Company, New York (1969).
7. Brockett, P. L.: Information Theoretic Approach to Actuarial Science: a Unification and Extension of Relevant Theory and Applications, TSA 43 (1991), pp. 73–113.
8. Brockett, P. L. és Cox, S.: Statistical Adjustment of Mortality Tables to Reflect Known Information, TSA 36(1984), pp. 63–71.
9. Charnes, A., Cooper, W. W. és Tyssedal, J.: Khinchine-Kullback-Leibler Estimation with Inequality Constraints, Math. Operationsforsch. Statist., Ser. Optimization 14 (1983), pp. 1–4.

INFORMATION THEORY IN ACTUARIAL SCIENCE

The paper deals with a branch of statistics in which optimization plays an important role. It presents the maximum entropy approach for incorporating prior information into the determination of probability distributions. First, the concept of 'information' and 'uncertainty' and their appropriate measures are introduced. Next, the maximum entropy methodology is presented in the form of a mathematical programming model, and duality statements are applied to develop the formula for its situation. Finally, the method is shown to be useful in several areas in actuarial modeling.