

Valószínűségi előrejelzések

Diplomamunka

Írta: Szamoránsky János

Matematikus szak

Témavezető:

Arató Miklós, egyetemi docens

Valószínűségelméleti és Statisztika Tanszék

Eötvös Loránd Tudományegyetem, Természettudományi Kar



Eötvös Loránd Tudományegyetem

Természettudományi Kar

2009

Tartalomjegyzék

1. Lényeges szkórok	4
1.1. Lényeges szkórok karakterizációja	4
1.1.1. Információmértékek és Bregman-divergenciák	6
1.2. Lényeges szkórok kategorikus változók esetén	6
1.2.1. Savage reprezentáció	7
1.2.2. Schervish reprezentáció	8
1.2.3. Példák	9
1.3. Lényeges szkórok sűrűségfüggvény-előrejelzések esetében	11
1.4. CRPS (Continuous Ranked Probability Score)	13
1.5. Az előrejelzői teljesítmény klasszikus mérőszámai	15
1.5.1. PMCC (Predictive Model Choice Criteria)	16
1.6. Energia szkór	17
1.7. Lényeges szkórok és negatív definit függvények kapcsolata	17
1.8. Szkórok intervallum- és kvantiliselőrejelzések esetében	19
1.8.1. Lényeges szkórok kvantilis előrejelzésekhez	19
1.8.2. Intervallumszkór	21
1.9. Optimális szkór becslés	22
2. Kalibráció és élesség	24
2.1. Kalibráció	24
2.1.1. Valószínűségi kalibráció becslése	27
2.1.2. Marginális kalibráció becslése	33
2.2. Élesség	34
2.2.1. Az élesség becslése	36
3. Bayesi megközelítés	37
3.1. Konjugált priorok	38
3.2. Hipotézisvizsgálat és Bayes faktor	41
3.2.1. Lindley paradoxona [9]	42
3.3. Bayes faktor az általános esetben	43
4. A logaritmikusszkór határeloszlása néhány esetben	45
4.1. Exponenciális modell	46
4.2. Normális modell	47
4.3. Lognormális modell	49
4.4. Gamma modell	50

Bevezetés

A statisztikai analízis legfontosabb céljai közé tartozik, hogy megbízható előrejelzéseket adjunk a jövőre nézve. Persze a jövő bizonytalansága miatt nem szorítkozhatunk csak determinisztikus előrejelzésekre. Gondolhatunk a jövőbeli eseményekre, mint valamilyen háttér-folyamat véletlenül bekövetkező kimeneteleire és vehetünk valószínűségi eloszlásokat egy ilyen jövőbeli esemény lehetséges kimenetelein, azaz külön-külön rendelhetünk valószínűségeket minden egyes lehetőséghez. Így az előrejelzésünk egy eloszlás lesz, melyet prediktív eloszlásnak nevezünk. A matematikai statisztika Bayes-bebecslésénél is tulajdonképpen az egész a posteriori eloszlást kaptuk bebecslésként, s ennek néztük különböző funkcionáljait (várható érték, medián) pontbecslésként. Az elmúlt két évtizedben az e fajta valószínűségi előrejelzések rengeteg alkalmazásra letek meteorológia előrejelzésektől (időjárás-¹, klímaelőrejelzések) kezdve különböző makroökonómiai prognózisokig, köszönhetően főleg az MCMC (Markov Chain Monte Carlo) módszerek különböző továbbfejlesztéseinek is, melyek a prediktív eloszlásból való Monte Carlo-mintavételek formájában kerülnek alkalmazásra. Először csak a diszkrét eset került a kutatók elé, de aztán olyan folytonos és diszkrét-folytonos változók, mint például hőmérséklet, csapadék, szélsébség, gdp, infláció kerültek előtérbe, amelyek segítették az általános elmélet kialakulását is. A valószínűségi előrejelzések ekkor prediktív sűrűség- ill. eloszlásfüggvények alakját öltik magukra. Továbbra is fennáll a kérdés: hogyan mérjük az ilyen valószínűségi előrejelzésekben meglevő bizonytalanságot?

A dolgozatom első részében bevezetésre kerülő szkórok alapvető szerepet játszanak a valószínűségi előrejelzések kiértékelésénél. Egy ilyen szkór az előrejelzéseken és megfigyeléseken értelmezett numerikus skála, melyekkel valószínűségi előrejelzések minőségét mérjük. Pontosabban egy szkór egy kétváltozós valós értékű leképezés, melynek egyik változója a prediktív eloszlás, a másik pedig a ω realizáció. Az így kapott értéket az előrejelzés jutalmának tekintjük. Gneiting és Raftery [2] cikke alapján megadom ezeknek egy igen fontos tulajdonságát, mely óvatosságra inti az előrejelzőt és őszinte előrejelzésekre biztatja. Minden ilyen szkór egy konvex függvényből ered, és mindegyiknek van egy információelméleti vonatkozása: mindegyikhez tartozik egy entrópia függvény, ill. (Bregman) divergencia. Példákat mutatok a gyakorlatban legtöbbször használt szkórokra, mint például a kvadratikusszkór, vagy a sokkal robusztusabb CRPS. A legegyszerűbb ilyen előrejelzés persze $p(\omega)$ lenne, ez azt az előrejelzőt helyezi előbbre, mely nagyobb valószínűséget csatolt a megfigyeléshez, viszont majd látni fogjuk, hogy ez nem fog megfelelni a megkövetelt tulajdonságnak. Ezért helyette a $LS(p, \omega) = \log p(\omega)$ képlettel definiált ún. logaritmikus szkórt használjuk majd. Bemutatom a megismert CRPS egy messzemenő általánosítását, mely Székely J. Gábor nevéhez fűződik, majd említést teszek a negatív definit függvények és a szkórok kapcsolatáról. E kapcsolat révén általános mintatér esetében tudunk értelmezni szkórokat. Ezután intervallum- és kvantiliselőrejelzések esetében használt szkórok egy általános alakját mutatom be. Mindezekmellett a bebecslélméletben is fontos szerep juthat a szkóroknak, ugyanis az általuk nyert hasznossági függvények segítségével igen sok különféle bebecslést kaphatunk, melyek között található néhány már eddig is ismert bebecslésfajta, mint pl. maximum likelihood vagy legkisebb négyzetes bebecslés. Az így kapott bebecslések a ML-bebecslés és az M-bebecslések

¹Valószínűségi időjárás-előrejelzések tipikusan *ensemble* prediktív rendszereken alapszanak, melyek az atmoszféra aktuális állapotának legjobb bebecslésében perturbációkat generálnak, és numerikus időjárás-előrejelzési modellek segítségével futtatják őket, s az így kapott eredményeket, mint a prediktív eloszlásból származó mintát kezelik valamilyen jövőbeli időjárási mennyiség előrejelzésére.

közé tehetők. Nagy előnyük, hogy a feladat természetétől függően mi választhatjuk meg melyik szkórhoz tartozó hasznossági függvénnyel² dolgozunk.

A dolgozatom második részében valószínűségi előrejelzések kalibrációjával foglalkozom. Gneiting, Raftery és mások [3]-ban a valószínűségi előrejelzések célját a következőképpen fogalmazza meg: maximalizálni az eloszlás élességét rögzített kalibráció mellett. A kalibráció az adatok konzisztenciáját jelenti, mennyire egyeztethető össze statisztikailag a prediktív eloszlásunk a megfigyelésekkel, így közös tulajdonsága az előrejelzésnek és a megfigyeléseknek. Míg az élesség a prediktív eloszlás koncentráltóságát fejezi ki, így csak magának az eloszlásnak a tulajdonsága. Minél koncentráltabb a prediktív eloszlás, annál élesebb az előrejelzés; és az élesebb előrejelzés a jobb, adott kalibráció mellett. A kalibráció különféle változatait definiálom és különböző eszközöket (például PIT, marginális kalibráció diagramm) ezek meglétének ellenőrzésére illetve becslésére. Mindezeket példákon keresztül és ábrákkal igyekszem ismertetni.

Ezután megvizsgálom, hogyan kezelik a valószínűségi előrejelzéseket bayesi környezetben. Ismertetem a konjugált priorokat, melyek adott likelihood mellett mindig ugyanabba az eloszláscsaládba tartozó posteriori eloszlásokat adnak, mint maguk. Egyszerűen belátható, hogy minden exponenciális eloszláscsaládba tartozó likelihood esetén mindig található hozzá szintén az exponenciális eloszláscsaládba tartozó konjugált prior. Bemutatom a Bayes faktort, először mint a hipotézisvizsgálat bayesi megfelelőjét, majd definiálom általános esetben is. Ismertetem ennek néhány hátrányát, mint például a priori hiperparaméterekre való érzékenységet, egyúttal megemlítem Lindley paradoxonát is. Ebben a fejezetben leginkább Walsh [10] cikkére támaszkodtam.

Majd a dolgozat utolsó részében az egyik leggyakrabban használt szkór, a logaritmikus szkór határeloszlását vizsgálom néhány ismert modell esetén. Ezután egy példán keresztül bemutatom, miért hasznos a lényeges szkórokat kitüntetű tulajdonság, majd az általános esetben is megmutatom, hogy ez teszi lehetővé, hogy 1-hez tartó valószínűséggel mindig a helyes modellt fogjuk kiválasztani, ha a lényeges szkórokat használjuk. Ezután egy saját példán keresztül is bemutatom a Bayes faktor a priori hiperparaméterekre vonatkozó érzékenységet.

²Vagy negatív orientációval veszteséggfüggvénnyel.

1. fejezet

Lényeges szkórok

Az első fejezetben először bevezetem a jelöléseket, megadom a lényeges szkórok egy osztályának, a reguláris szkóroknak a karakterizációját, majd ismertetem a megfelelő információelméleti vonatkozásokat, illetve Bregman divergenciákat. Ezután a folytonos előrejelzések esetével foglalkozom, bemutatom a CRPS-t, majd ennek messzemennő általánosításait; megvizsgálom a lényeges szkórok és a negatív definit függvények kapcsolatát, végül az intervallum- és kvantiliselőrejelzések esetével foglalkozom, illetve említést teszek az optimális szkór becslésről, mely magába foglal sok eddig megismert becsléstípust.

1.1. Lényeges szkórok karakterizációja

Valószínűségi előrejelzések vizsgálatához először is meg kell adni egy Ω mintateret, illetve az Ω részhalmazaiából álló \mathcal{A} mérhető halmazok σ -algebráját. Az így kapott (Ω, \mathcal{A}) mérhető téren legyen továbbá adott valószínűségi mértékek egy \mathcal{P} konvex családja.

Ekkor egy Ω -n értelmezett kiterjesztett valós értékű függvényt \mathcal{P} -kváziintegrálhatónak nevezünk, ha mérhető \mathcal{A} -ra nézve és kváziintegrálható minden $P \in \mathcal{P}$ szerint. Kváziintegrálhatóság alatt azt értve, hogy az integrál létezik, az értéke lehet véges vagy végtelen.

Valószínűségi előrejelzésnek nevezzük \mathcal{P} egy P elemét.

Egy $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ kiterjesztett valós értékű függvényt *szkór*nak vagy *pontozó szabálynak* nevezünk, ha $S(P, \cdot)$ \mathcal{P} -kváziintegrálható minden $P \in \mathcal{P}$ esetén. Egy ilyen szkór úgy interpretálható, hogy ha egy előrejelző a $P \in \mathcal{P}$ valószínűségi előrejelzést adja és ω realizálódik, akkor az előrejelző jutalma $S(P, \omega)$.

Bevezetem még a következő jelölést a Q szerinti *várható szkórra* a P előrejelzés mellett:

$$S(P, Q) = \int S(P, \omega) dQ(\omega) \tag{1.1}$$

Ezt felfoghatjuk mint a P előrejelzésünk várható jutalmát, ha a valódi háttéreloszlás Q . Fontos lenne olyan szkórokat készíteni melyek "őszinte" előrejelzésekre bíztatnak. Minél "közelebb" van az előrejelzésünk a valódi háttéreloszláshoz, annál jobb munkát végeztünk és nagyobb jutalmat érdemlünk. Tehát az eloszlások között kellene valamilyen távolságfogalmat definiálni.

A következőkben egy S szkór igen fontos tulajdonságát definiálom, lényegében csak az ilyen szkórokkal kell foglalkoznunk. Egy S szkórt *lényegesnek*¹ nevezünk a \mathcal{P} osztályon, ha

$$S(Q, Q) \geq S(P, Q) \quad \forall P, Q \in \mathcal{P} \quad (1.2)$$

Azaz bármely háttéreloszlás esetén a várható szkór maximális, hogyha ezt a valódi eloszlást (a lehető legjobbat) adjuk előrejelzésként. Ha a maximum csak ekkor éretik el, és minden más esetben szigorú egyenlőtlenség áll, akkor az S -t *szigorúan lényegesnek* nevezzük. Azt, hogy miért ez a megfelelő definíció, majd a későbbiekben látni fogjuk. Világos, hogy (szigorúan) lényeges szkórok és \mathcal{P} -integrálható függvények véges összege (szigorúan) lényeges.²

Definíció: Egy $G : \mathcal{P} \rightarrow \mathbb{R}$ függvény konvex, ha $\forall \lambda \in (0, 1), \forall P_0, P_1 \in \mathcal{P}$ esetén

$$G((1 - \lambda)P_0 + \lambda P_1) \leq (1 - \lambda)G(P_0) + \lambda G(P_1) \quad (1.3)$$

A G szigorúan konvex, ha egyenlőség csak $P_0 = P_1$ esetben lehetséges.

Definíció: Egy $G^*(P, \cdot) : \Omega \rightarrow \overline{\mathbb{R}}$ függvényt G *szubtangensének* nevezzük a $P \in \mathcal{P}$ pontban, ha integrálható P szerint, kváziintegrálható a többi $Q \in \mathcal{P}$ szerint és

$$G(Q) \geq G(P) + \int G^*(P, \omega) d(Q - P)(\omega) \quad \forall Q \in \mathcal{P} \quad (1.4)$$

Definíció: Egy $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ szkór *reguláris* \mathcal{P} -n, ha $\forall P, Q \in \mathcal{P}$ esetén teljesül: $S(P, P) \in \mathbb{R}$ és $S(P, Q) < \infty$.

Tétel: Az $S : \mathcal{P} \times \Omega \rightarrow \overline{\mathbb{R}}$ *reguláris szkór akkor és csak akkor (szigorúan) lényeges* \mathcal{P} -n, ha létezik egy (szigorúan) konvex, valós értékű G függvény \mathcal{P} -n, melyre

$$S(P, \omega) = G(P) - \int G^*(P, \omega) dP(\omega) + G^*(P, \omega) \quad P \in \mathcal{P}, \omega \in \Omega \quad (1.5)$$

ahol $G^*(P, \cdot) : \Omega \rightarrow \overline{\mathbb{R}}$ G egy szubtangense $P \in \mathcal{P}$ -ben.³

Bizonyítás: Ha az S reguláris szkór a fenti alakú, akkor a (1.4) szubtangens egyenlőtlenségből azonnal következik a lényegességet definiáló egyenlőtlenség. Fordítva legyen S egy reguláris, lényeges szkór. Defináljuk $G : \mathcal{P} \rightarrow \mathbb{R}$ -t a következőképpen: $G(P) = S(P, P) = \sup_{Q \in \mathcal{P}} S(Q, P)$, ami konvex függvények pontonkénti szuprémuma, így maga is konvex. Továbbá a szubtangens egyenlőtlenség is teljesül $G^*(P, \omega) = S(P, \omega)$ választással. Ebből következik az előállítás, és bizonyítja a lényegességet. Szigorú egyenlőtlenség ekvivalens azzal, hogy G P -beli szubtangense nem egyenlő G Q -beli szubtangensével, és ez ekvivalens G konvexitásával \mathcal{P} -n. \square

Kicsit másképp fogalmazva: az S reguláris szkór pontosan akkor (szigorúan) lényeges \mathcal{P} -n, ha a $G(P) = S(P, P)$ várható szkór függvény (szigorúan) konvex \mathcal{P} -n és $S(P, \omega)$ G egy szubtangense P -ben minden $P \in \mathcal{P}$ -re.

¹Paraméteres környezetben Lehmann és Casella (1998) az (1.2) definiáló tulajdonságra, mint kockázat torzítatlanságára (risk unbiasedness) hivatkozik. [2]

²Az angol irodalomban a *proper scoring rules* néven ismertek, nem találtam a magyar irodalomban fordítást a *proper* kifejezésre, így a magam általi *lényeges* fordítást használom, remélem elég kifejező.

³Ekkor persze $G(P) = \int S(P, \omega) dP(\omega) = S(P, P) = \sup_{Q \in \mathcal{P}} S(Q, P)$.

1.1.1. Információmértékek és Bregman-divergenciák

Legyen S egy lényeges szkór \mathcal{P} -n, ekkor a

$$G(P) = S(P, P) = \sup_{Q \in \mathcal{P}} S(Q, P) \quad (1.6)$$

várható szkór függvényt az S szkórhoz tartozó (*általánosított*) *entrópiának* nevezzük. Ez a maximális elérhető hasznosság és a sima *entrópia* kifejezés is használatos. Ha az S szkór reguláris és lényeges, akkor

$$d(P, Q) = S(P, P) - S(Q, P) \quad P, Q \in \mathcal{P} \quad (1.7)$$

az S -hez tartozó *divergencia függvény*. Ez nemnegatív, és ha az S szkór szigorúan lényeges, akkor pontosan akkor 0, ha $P = Q$. Tehát a várható szkór maximalizálása egyenértékű a divergencia minimalizálásával.⁴ Bizonyos esetekben előfordulhat, hogy valamilyen korlátozások miatt csak adott típusú eloszlások jöhetnek szóba előrejelzésként, így ezek közül kell kiválasztani a legjobbat, ekkor ezek között kell minimalizálni a divergenciát. Például ha egy paraméteres eloszláscsaládot akarunk illeszteni az adatsorra, akkor ezen a családon belül minimalizálhatjuk a divergenciát és így becslést kapunk a paraméterre. Ha a mintatér véges, és az entrópia függvény elegendően sima, akkor a divergencia függvény a *Bregman divergencia*, melynek fontos szerep jut az optimalizációelméletben. A *Bregman távolság* kifejezést is használják habár $d(P, Q) = d(Q, P)$ nem feltétlen teljesül minden esetben.

Érdekes probléma olyan feltételeket találni, mely mellett egy d divergenciához megadható egy S szkór úgy, hogy d a megfelelő szkór-divergencia. Ekkor persze egy új út nyílik előttünk különböző szkórok definiálásához. Szimmetrikus divergencia esetében Savage ([2], 1971) ad szükséges feltételt: ha P és Q ugyanarra az eseményre koncentrált indikátorszámok a megfelelő p és $q \in [0, 1]$ valószínűségekkel, akkor $d(P, Q)$ $(p-q)^2$ lineáris függvényévé redukálódik. Friedman és Nau ([2], 1985) egy lazább kapcsolatot vizsgált lényeges szkórok és valószínűségi mértékek távolságai közt. Vizsgálataikat valódi metrikákra korlátozták, melyek szimmetrikusak és kiegyenlítik a háromszög-egyenlőtlenséget is. Egy S szkórt effektívnek neveznek egy d metrikával, ha

$$S(P_1, Q) \geq S(P_2, Q) \iff d(P_1, Q) \leq d(P_2, Q) \quad (1.8)$$

Egy d metrikát pedig ko-effektívnek, ha létezik olyan S lényeges szkór, mely effektív d -vel. Bebizonyították, hogy az l_1 , l_∞ és a Hellinger távolság⁵ az abszolút folytonos valószínűségi mértékek terén nem ko-effektívek.

1.2. Lényeges szkórok kategórikus változók esetén

Most az előző két speciális esetét vizsgálom, és véges mintatér esetén Savage (1971, [2]) és Schervish (1989, [2]) két reprezentációját mutatom be, melyek kategórikus és bináris változók esetén karakterizálja a (szigorúan) lényeges szkórokat, majd néhány példát láthatunk lényeges szkórokra.

⁴Ez felelhet meg a két eloszlás "távolságának", bár nyilvánvaló, hogy matematikai értelemben nem metrikáról van szó, például egyáltalán nem biztos, hogy szimmetrikus.

⁵Hellinger távolság: $d(P, Q) = \int (\sqrt{p(x)} - \sqrt{q(x)})^2$, ahol p és q a megfelelő sűrűségfüggvények.

1.2.1. Savage reprezentáció

Először tehát kategórikus vagy más néven kvalitatív változók előrejelzéseivel foglalkozom. Ezek véges sok értéket felvevő diszkrét valószínűségi változók, melyeknél az értékek nem nagyságrendet tükröznek, hanem a változó olyan kategórikus értékeit hívatottak jelölni, mint például szín vagy területi felosztás. A bináris változók ennek azt a speciális esetét adják, mikor két elemi eseményünk van, és dönteni szeretnénk, hogy melyik következik be közülük. Tehát akkor a mintatér $\Omega = \{1, \dots, m\}$ véges sok, egymást kizáró eseményből áll, és egy valószínűségi előrejelzés egy (p_1, \dots, p_m) valószínűségi vektor. Az eddigi jelöléseket használva az előrejelzések konvex osztálya

$$\mathcal{P} = \mathcal{P}_m = \left\{ p = (p_1, \dots, p_m) : p_1, \dots, p_m \geq 0, p_1 + \dots + p_m = 1 \right\}$$

Ekkor egy S szkór megadása ekvivalens a következő m függvény megadásával:

$$S(\cdot, i) : \mathcal{P}_m \rightarrow \overline{\mathbb{R}} \quad i = 1, \dots, m$$

Vagyis, ha egy előrejelző a p valószínűségi vektort adja és az i esemény következik be, akkor az ő jutalma $S(p, i)$. A következő tétel az előző pontban ismertett (1.5) speciális esete, és Savage reprezentációját adja a lényeges szkóroknak véges mintatér esetén.

Előbb definiálom az előző pont fogalmait erre a speciális esetre:

Definíció: Legyen $G : \mathcal{P}_m \rightarrow \mathbb{R}$ konvex függvény. Ekkor egy $G'(p) = (G'_1(p), \dots, G'_m(p))$ vektort⁶ G szubgradiensének nevezünk $p \in \mathcal{P}_m$ -ben, ha

$$G(q) \geq G(p) + \langle G'(p), q - p \rangle \quad \forall q \in \mathcal{P}_m \quad (1.9)$$

ahol $\langle \cdot, \cdot \rangle$ jelöli a (standard) skaláris szorzatot.

Definíció: Egy S szkór kategórikus változókra *reguláris*, ha $S(\cdot, i)$ valós értékű, kivéve esetleg $S(p, i) = -\infty$, ha $p_i = 0$.

Tétel (Savage reprezentáció)

Egy S kategórikus változókra vonatkozó reguláris szkór akkor és csak akkor (szigorúan) lényeges, ha

$$S(p, i) = G(p) - \langle G'(p), p \rangle + G'_i(p) \quad i = 1, \dots, m \quad (1.10)$$

ahol $G : \mathcal{P}_m \rightarrow \mathbb{R}$ egy (szigorúan) konvex függvény és $G'(p)$ G egy szubgradiense p -ben $\forall p \in \mathcal{P}_m$ esetén.

Azaz egy S reguláris szkór akkor és csak akkor (szigorúan) lényeges, ha a $G(p) = S(p, p)$ várható szkór függvény (szigorúan) konvex \mathcal{P}_m -n és az $S(p, i)$ ($i = 1, \dots, m$) komponensekből alkotott vektor G egy szubgradiense p -ben.

Tehát ez eddigiek szerint minden korlátos (szigorúan) konvex G függvény \mathcal{P}_m -en generál egy reguláris, (szigorúan) lényeges szkórt. Ez a G függvény lesz a (1.6) várható szkór függvény, információérték vagy entrópiafüggvény, a megfelelő divergencia pedig (1.7) Bregman távolság.

⁶Tegyük fel, hogy $G'_i(p)$ valós értékű, kivéve a $G'_i(p) = -\infty$, ha $p_i = 0$ esetet.

1.2.2. Schervish reprezentáció

Mint már említettem a klasszikus eldöntendő kérdések témaköre külön tárgyalást érdemel. Az eseménytér most két elemből áll, ahol az 1 felel meg az *igennek*: $\Omega = \{0, 1\}$. Egy valószínűségi előrejelzés legyen $p \in [0, 1]$ az *igen* valószínűsége.⁷ Ebben az esetben egy S szkór a következő két függvénnyel adható meg:

$$S(\cdot, 1) : [0, 1] \rightarrow \overline{\mathbb{R}}, \quad S(\cdot, 0) : [0, 1] \rightarrow \overline{\mathbb{R}}$$

Ezek azt adják meg, hogy mekkora a jutalmunk hogyha p valószínűséggel jeleztük előre az eseményt és az bekövetkezik ($S(p, 1)$), ill. nem következik be ($S(p, 0)$).

Definíció: Az S szkór *reguláris*, ha $S(\cdot, 1)$ és $S(\cdot, 0)$ végesek, kivéve esetleg az $S(0, 1) = -\infty$ vagy $S(1, 0) = -\infty$ eseteket.⁸

A (1.5) tétel erre az esetre vonatkozó variánsa szerint minden reguláris, (szigorúan) lényeges szkór a következő alakú:

$$S(p, 1) = G(p) + (1 - p)G'(p), \quad S(p, 0) = G(p) - pG'(p) \quad (1.11)$$

ahol $G : [0, 1] \rightarrow \mathbb{R}$ (szigorúan) konvex függvény és $G'(p)$ G egy szubgradiense p -ben, a következő értelemben:

$$G(q) \geq G(p) + G'(p)(q - p) \quad \forall q \in [0, 1] \quad (1.12)$$

A $G'(p)$ szubgradiens valós értékű, kivéve a $G'(0) = -\infty$ és a $G'(1) = -\infty$ lehetséges eseteket. Ha G differenciálható a $p \in (0, 1)$ belső pontban, akkor a $G'(p)$ értéke egyértelmű és megegyezik a derivált értékével.⁹

A Savage reprezentációból igen érdekes tulajdonságait nyerhetjük a reguláris, (szigorúan) lényeges szkóroknak. Például a következő

$$S(p, 1) = \lim_{q \rightarrow 1} G(q) - \int_p^1 (G'(q) - G'(p))dq \quad p \in (0, 1) \quad (1.13)$$

előállításból leolvasható (mivel $G(p)$ (szigorú) konvexitása miatt $G'(p)$ (szigorúan) monoton nő), hogy $S(p, 1)$ szintén (szigorúan) monoton nő. Hasonlóan $S(p, 0)$ (szigorúan) monoton csökken, ahogy azt tőlük előzetesen is elvárnánk. Ugyanis minél nagyobb valószínűséggel jeleztünk előre egy eseményt, ami aztán bekövetkezett, annál nagyobb jutalomra számítunk. Viszont ha nem következett be az esemény, akkor a jutalmunk egyre csökken, minél nagyobb valószínűséggel jeleztük előre azt.

Ennek felhasználásával bizonyítható a következő tétel, mely a bináris előrejelzésekre vonatkozó reguláris lényeges szkórokat karakterizálja:

⁷Különbség az előző részben tárgyaltakhoz képest, hogy most $\mathcal{P} = [0, 1]$, míg ott $\mathcal{P}_2 = \{(p_1, p_2) \in \mathbb{R}^2 : p_1 \in [0, 1], p_2 = 1 - p_1\}$ volt.

⁸Ezek azok az esetek, amikor az előrejelzés szerint az esemény 1 valószínűséggel be fog következni, ám nem következik be, illetve fordítva: az előrejelzés szerint 1 valószínűséggel nem fog következni az esemény, mégis bekövetkezik.

⁹Egy szubgradiens lényegében egy támaszhipersíknak felel meg.

Tétel (Schervish reprezentáció)

Legyen S egy reguláris szkór. Ez akkor és csak akkor lényeges, és $S(0, 1) = \lim_{p \rightarrow 0} S(p, 1)$, $S(0, 0) = \lim_{p \rightarrow 0} S(p, 0)$, illetve $S(p, 0)$ és $S(p, 1)$ balról folytonosak, ha létezik egy olyan ν mérték $(0, 1)$ -en, melyre

$$S(p, 1) = S(1, 1) - \int_{[p, 1)} (1 - q)\nu(dq) \quad \forall p \in [0, 1] \quad (1.14)$$

$$S(p, 0) = S(0, 0) - \int_{[0, p)} q\nu(dq) \quad \forall p \in [0, 1] \quad (1.15)$$

Az S szkór akkor és csak akkor szigorúan lényeges, ha ν pozitív súlyt rendel minden nyílt intervallumhoz.

Bizonyítás: (vázlat) Tegyük fel, hogy S kielégíti a tétel feltételeit. $S(p, 1)$ (1.14) alakjához nézzük a (1.13) reprezentációt és mivel $G'(p)$ monoton növekszik, ezért azonosítsuk egy ν mérték (balról folytonos) eloszlásfüggvényével, és alkalmazzunk parciális integrálást. $S(p, 0)$ előállítására analóg módon történik. A másik irányhoz fordítsuk meg a gondolatmenetet. A szigorú lényegességre vonatkozó állítás a szigorúan konvex függvények ismert tulajdonságaiból következik. \square

Ha a G várható szkór függvény elegendően sima, akkor $\nu(dq)$ -nak létezik Lebesgue-sűrűségfüggvénye $-G''(q)$. Például a Brier szkór esetében $G(p) = 2p(1 - p)$, így a hozzátartozó ν mérték az egyenletes. Míg a logaritmikus szkór a Shannon entrópiából származik, és így $G(p) = p \log p + (1 - p) \log(1 - p)$, a ν ekkor egy nem-véges mérték $(q(1 - q))^{-1}$ Lebesgue-sűrűségfüggvénnyel.

1.2.3. Példák

Az eddigiek után lássuk néhány példát is lényeges szkórokra. Mindegyiknek létezik bináris változata is, de ezt nem részletezem mindenhol, ugyanis ezt megkaphatjuk, hogy ha a képletet két eseményre írjuk fel.

• Kvadratikus szkór (vagy Brier szkór)¹⁰

Legyen $G(p)$ a következő konvex függvény \mathcal{P}_m -en: $G(p) = \sum_{j=1}^m p_j^2 - 1$

Ekkor a (1.10) előállítás a következő ún. kvadratikus szkórt adja:

$$S(p, i) = 2p_i - \sum_{j=1}^m p_j^2 - 1 = - \sum_{j=1}^m (\delta_{ij} - p_j)^2 \quad (1.16)$$

ahol δ_{ij} a Kronecker-delta:

$$\delta_{ij} = \begin{cases} 1 & \text{ha } i = j \\ 0 & \text{ha } i \neq j \end{cases}$$

A megfelelő Bregman divergencia a négyzetes euklideszi távolság lesz:

$$d(p, q) = \sum_{j=1}^m (p_j - q_j)^2 \quad (1.17)$$

¹⁰Ezt az igen elterjedt szkórt Brier (1950) vezette be, innen is a Brier szkór elnevezés.

- **(Pszseudo)szférikus szkór**

Legyen $\alpha > 1$ és nézzük a következő általánosított entrópiafüggvényt: $G(p) = (\sum_{j=1}^m p_j^\alpha)^{1/\alpha}$.
Az ehhez tartozó pszeudoszférikus szkór:

$$S(p, i) = \frac{p_i^{\alpha-1}}{(\sum_{j=1}^m p_j^\alpha)^{(\alpha-1)/\alpha}} \quad (1.18)$$

A tradicionális sférikus szkórt az $\alpha = 2$ esetben kapjuk:

$$S(p, i) = \frac{p_i}{\|p\|_2} \quad (1.19)$$

A megfelelő Bregman divergencia:

$$d(p, q) = \left(\sum_{j=1}^m p_j^\alpha\right)^{1/\alpha} - \frac{\sum_{j=1}^m p_j q_j^{\alpha-1}}{(\sum_{j=1}^m q_j^\alpha)^{(\alpha-1)/\alpha}} \quad (1.20)$$

- **Logaritmikus szkór**

A negatív Shannon entrópia, $G(p) = \sum_{j=1}^m p_j \log p_j$ szintén meghatároz egy lényeges szkórt, melyet logaritmikus szkórnak nevezünk:

$$S(p, i) = \log p_i \quad (1.21)$$

A hozzátartozó Bregman divergencia a Kullback-Leibler divergencia:

$$d(p||q) = \sum_{j=1}^m p_j \log(p_j/q_j) \quad (1.22)$$

Ez a klasszikus szkór egészen Good ([13]) munkásságáig vezethető vissza, a gyakorlatban az egyik legtöbbet alkalmazott szkórról van szó.

- **0-1 szkór**

A következő szkór abban az esetben jutalmaz egy valószínűségi előrejelzést, ha annak a módusza realizálódik. Ha több módusz is van, akkor a jutalom arányosan oszlik el ezek között:

$$S(p, i) = \begin{cases} \frac{1}{|M(p)|} & \text{ha } i \in M(p) \\ 0 & \text{különben} \end{cases} \quad (1.23)$$

ahol $M(p) = \{i : p_i = \max_{j=1\dots m} p_j\}$

A szkórnak megfelelő általánosított entrópia $G(p) = \max_{j=1\dots m} p_j$, a divergencia függvény pedig:

$$d(p, q) = \max_{j=1\dots m} p_j - \frac{\sum_{j \in M(q)} p_j}{|M(q)|} \quad (1.24)$$

Ez nem lesz Bregman divergencia, mivel az entrópia függvény se nem differenciálható, se nem szigorúan konvex.

A megismert szkórok mindegyike *szimmetrikus* a következő értelemben:

$$S((p_1, \dots, p_m), i) = S((p_{\pi_1}, \dots, p_{\pi_m}), \pi_i) \quad \forall p \in \mathcal{P}_m, \forall \pi \in \mathcal{S}_m, \forall i$$

Winkler ([14], 1994) szerint a szimmetrikus szkórok nem mindig díjazták megfelelően az előrejelzői képességeket, és aszimmetrikus szkórok használatára bíztat. Aszimmetrikus (szigorúan) lényeges szkórokat olyan G (szigorúan) konvex entrópia függvényekből kaphatunk, melyek nem invariánsak a koordinátapermutációra.

1.3. Lényeges szkórok sűrűségfüggvény-előrejelzések esetében

Legyen μ egy σ -véges mérték a (Ω, \mathcal{A}) mérhető téren. Jelölje \mathcal{L}_α ($\alpha > 1$) valószínűségi mértékek azon osztályát, melyek abszolút folytonosak μ -re és a sűrűségfüggvényeikre az

$$\|p\|_\alpha = \left(\int p(\omega)^\alpha \mu(d\omega) \right)^{1/\alpha} \quad (1.25)$$

integrál véges. Egy $P \in \mathcal{L}_\alpha$ valószínűségi előrejelzést a p sűrűségfüggvényével adunk meg és p -t *prediktív sűrűségfüggvénynek* (*sűrűségfüggvény-előrejelzés*) nevezzük.¹¹ Ebben az esetben is érvényes a reguláris, lényeges szkórokra az (1.5) előállítás, azaz az S reguláris szkór pontosan akkor (szigorúan) lényeges \mathcal{P} -n, ha a $G(P) = S(P, P)$ várható szkór függvény (szigorúan) konvex \mathcal{P} -n és $S(P, \omega)$ G egy szubtangense P -ben minden $P \in \mathcal{P}$ -re.

Most néhány a diszkrét esetben látott szkór abszolút folytonos megfelelőjét tekintem át.

Példák

- Az első szkór a kvadratikus (Brier) szkór volt, melynek folytonos alakja:

$$QS(p, \omega) = 2p(\omega) - \|p\|_2^2 \quad (1.26)$$

Állítás: *Ez szigorúan lényeges az \mathcal{L}_2 osztályon.*

Bizonyítás: Ehhez a $d(p, q) = QS(p, p) - QS(q, p) \geq 0$ feltételt kell leellenőrizni tetszőleges p és q \mathcal{L}_2 -beli sűrűségfüggvényekre, egyenlőtlenséget csak $p = q$ μ -m.m. esetben megengedve. Kiszámolva a divergenciát egyszerűen adódik: $d(p, q) = \|p - q\|_2^2 \geq 0$ és egyenlőség is csak a megengedett esetben lehetséges. \square

Láthatjuk, hogy a divergencia most is a négyzetes euklideszi távolság, míg az entrópia ebben az esetben $G(p) = \|p\|_2^2$, a p prediktív sűrűségfüggvény \mathcal{L}_2 -normanégyzete.

- Good ([13]) javasolta a folytonos pszeudoszférikus szkór használatát:

$$PseudoS_\alpha(p, \omega) = \frac{p(\omega)^{\alpha-1}}{\|p\|_\alpha^{\alpha-1}} \quad (1.27)$$

Szintén szigorúan lényeges az \mathcal{L}_α osztályon, ugyanis a $G(p) = \|p\|_\alpha$ entrópia szigorú konvexitása egyszerű következménye a Minkowski-egyenlőtlenségnek, míg a

$$d(p, q) = \|p\|_\alpha - \frac{\int q^{\alpha-1}(\omega)p(\omega)\mu(d\omega)}{\|q\|_\alpha^{\alpha-1}} \quad (1.28)$$

divergencia nemnegativitása a Hölder-egyenlőtlenségből adódik.

Az $\alpha = 2$ esetben a klasszikus szférikus szkórt kapjuk:

$$PseudoS_2(p, \omega) = \frac{p(\omega)}{\|p\|_2} \quad (1.29)$$

¹¹Ezek csak μ -m.m. vannak meghatározva.

- A következő a logaritmikus szkór folytonos esete:

$$\text{Log}S(p, \omega) = \log p(\omega) \quad (1.30)$$

mely megkapható megfelelően paraméterezett pszeudoszférikus szkór $\alpha \rightarrow 1$ határeseteként is. Bevezetését szintén Good ([13]) javasolta, és azóta nagyon széles körben elterjedt.¹²

A hozzá tartozó várható szkór függvény a negatív Shannon entrópia,

$$G(p) = \int p(x) \log p(x) \mu(dx) \quad (1.31)$$

míg a divergencia függvény a klasszikus Kullback-Leibler divergencia:

$$d(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} \mu(dx) \quad (1.32)$$

Állítás: *A logaritmikus szkór szigorúan lényeges az \mathcal{L}_1 osztályon.*

Bizonyítás: Ehhez a következőnek kell teljesülnie tetszőleges p és q sűrűségfüggvényekre:

$$\int p(x) \log p(x) \mu(dx) \geq \int p(x) \log q(x) \mu(dx)$$

és egyenlőség csak $p = q$ μ -m.m. esetben lehetséges.

$$\begin{aligned} \int p(x) \log \frac{q(x)}{p(x)} \mu(dx) &= E_p \left(\log \frac{q(X)}{p(X)} \right) \leq \\ &\leq \log E_p \left(\frac{q(X)}{p(X)} \right) = \log \int \frac{q(x)}{p(x)} p(x) \mu(dx) = 0 \end{aligned}$$

A Jensen-egyenlőtlenség alapján, és egyenlőség csak $\frac{q(x)}{p(x)} = \text{konstans}$ esetben lehetséges, de ekkor $p(x) = q(x)$ μ -m.m. \square

Láthatjuk, hogy a logaritmikus szkór csak a megfigyelt ω -n keresztül függ a prediktív sűrűségfüggvénytől, míg az előző kettő emellett a prediktív sűrűségfüggvény valamilyen \mathcal{L}_α -normáját is tartalmazta. Az előbbi tulajdonsággal rendelkező szkóroknak külön elnevezése is van:

Definíció: Egy olyan S szkórt, mely csak magán a megfigyelt ω -n keresztül függ a p prediktív sűrűségfüggvénytől *lokális szkór*nak nevezünk.

Bizonyos regularitási feltételek mellett megmutatható, hogy minden lényeges lokális szkór a következő alakú:

$$S(p, \omega) = a \log p(\omega) + f(\omega) \quad (1.33)$$

ahol $a \geq 0$ és $f : \Omega \rightarrow \mathbb{R}$ tetszőleges függvény. Azaz a lényeges lokális szkórok osztálya a logaritmikus szkór efféle transzformáltjaiból áll. Így a bevezetőben már említett a $\text{Lin}S(p, \omega) = p(\omega)$ formulával definiált *lineáris szkór* sem lesz lényeges, akármilyen intuitív is lenne a használata. Erről a következő egyszerű példán is megbizonyosodhatunk.

¹²Más néven is használatos, mint például *predictive deviance* vagy *ignorance score*.

Jelölje φ a standard normális sűrűségfüggvényt, míg u a $(-\epsilon, \epsilon)$ -on egyenletes eloszlását. Ekkor $\epsilon < \sqrt{\log 2}$ esetén

$$\text{LinS}(u, \varphi) = \frac{1}{\sqrt{2\pi}} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} e^{-x^2/2} dx > \frac{1}{2\pi^{1/2}} = \text{LinS}(\varphi, \varphi) \quad (1.34)$$

ellentmondásban a lényegesség (1.2) definíciójával. Valójában a lineáris szkór hajlamos túlbecslésre a valódi sűrűségfüggvény csúcsainál.

Wilson, Burrows and Lanzinger ([2], 1999) *valószínűségi szkórja* kiintegrálja a sűrűségfüggvényt a megfigyelt ω egy kis környezetében:

$$\text{VS}_{\epsilon}(P, \omega) = \int_{\omega-\epsilon}^{\omega+\epsilon} p(x) dx \quad (1.35)$$

de ez sem lesz lényeges.

A logaritmusos szkór talán egyetlen problémája, hogy nem elég robusztus, a kis valószínűségű megfigyeléseket igen erősen bünteti, és így igen érzékeny az extrém értékekre. Weigend és Shi ([2], 2000) ezért a logaritmusos szkór használatakor trimmelt átlagszkór számítását javasolja. A következő szkór sokkal robusztusabb lesz és nem is annyira érzékeny az extrém értékek esetén.

1.4. CRPS (Continuous Ranked Probability Score)

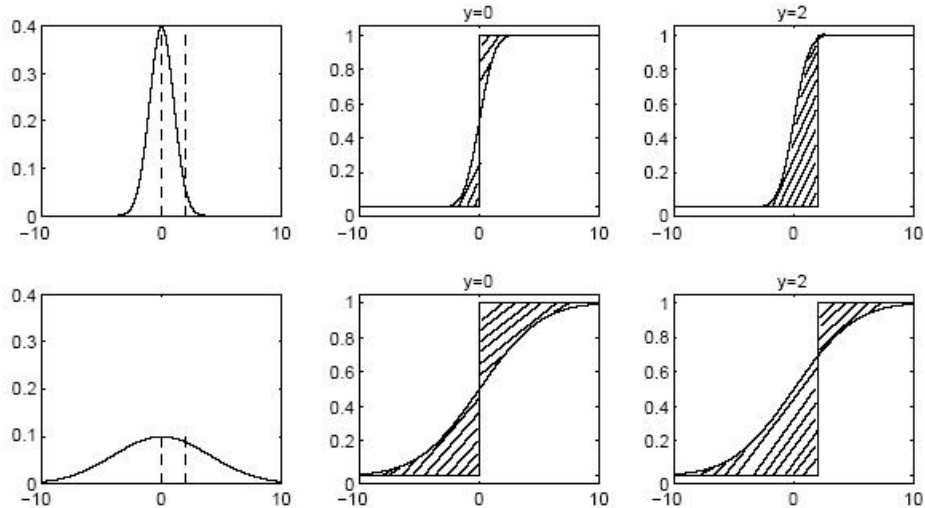
Ezidáig csak diszkrét vagy abszolút folytonos, sűrűségfüggvénnyel rendelkező előrejelzésekkel foglalkoztam. Utóbbi esetben maga a sűrűségfüggvény volt az előrejelzés. Ez egy nem praktikus megszorítás, mivel nem minden esetben létezik sűrűségfüggvény. Például, ha csapadékmennyiséget akarunk előrejelezni, akkor szükségszerűen pozitív súly kerül a 0-ba is. Persze le lehetne cserélni a Lebesgue-mértéket egy keverék domináló mértékre, de praktikusabb, ha eloszlásfüggvényeken értelmezett szkórokkal is dolgozunk. Továbbá az eddig tárgyalt szkórok nem nagyon érzékenyek a távolságra, abban az értelemben, hogy semmi se biztosítja, hogy a megfigyelt érték környezetére (a megfigyelt érték nélkül) is nagy súlyt helyező előrejelzéseknek kedveznek. Pedig ez igen kívánatos tulajdonság, ha a prediktív eloszlásaink többszűcsűak. Ezért lépünk egy lépcsőfokkal feljebb, hogy egy kalap alatt kezelhessük a két esetet, és vizsgáljuk az eloszlásfüggvények előrejelzésének esetét is. Ehhez jelölje \mathcal{P} Borel-mértékek egy családját \mathbb{R} -n. Egy valószínűségi előrejelzést, azaz \mathcal{P} egy elemét identifkáljuk F eloszlásfüggvényével.

Egy ilyen előrejelzés kiértékelésére definiálható a következő szkór:

$$\text{CRPS}(F, x) = - \int_{-\infty}^{\infty} (F(y) - \mathbf{1}\{y \geq x\})^2 dy \quad (1.36)$$

Ez a prediktív eloszlásfüggvénynek és az x megfigyelés empirikus eloszlásfüggvényének integrált négyzetes különbsége. Vegyük észre, hogy az integrál mögötti kifejezéshez hasonlóval már találkoztunk a Brier szkór definíciójánál. Vegyünk egy F eloszlásfüggvényű X -t, ekkor valójában az $\{X < y\}$ eseményre felírt bináris Brier szkór van az integrál mögött. Tehát a CRPS felfogható úgy is, mint az y küszöbértékekhez tartozó Brier szkórok egy integrálja.

A CRPS egy grafikus illusztrációja látható a következő ábrán normális prediktív eloszlás esetén. A bal oldalon látható a 0 várható értékű normális sűrűségfüggvény 1 ill. 4 szórnégyzettel, mellette pedig a megfelelő eloszlásfüggvény és az empirikus eloszlásfüggvény különbsége az $y = 0$ és az $y = 2$ megfigyelések esetén. Ezek az ábrák mutatják, hogy a CRPS jutalmazza az éles előrejelzéseket, de figyelembe veszi azt is, hogy a megfigyelés mennyire esik az eloszlás farokrészébe.



1.1. ábra. Normális és a tapasztalati eloszlásfüggvény különbsége [6]

Eleinte elég nehézkesnek tűnt ennek a szkórnak az alkalmazása analitikus kifejezés hiányában, ám a következő előállítás sokat segített ezen (Székely, [12]):

$$CRPS(F, x) = \frac{1}{2} E_F |X - X'| - E_F |X - x| \quad (1.37)$$

ahol X és X' független változók F eloszlásfüggvénnyel és véges első momentummal. Például ha a prediktív eloszlásunk normális, akkor ez a következő alakot ölti:

$$CRPS(N(\mu, \sigma^2), x) = \sigma \left(\frac{1}{\sqrt{\pi}} - 2\varphi \left(\frac{x - \mu}{\sigma} \right) - \frac{x - \mu}{\sigma} \left(2\Phi \left(\frac{x - \mu}{\sigma} \right) - 1 \right) \right) \quad (1.38)$$

ahol φ és Φ a standard normális sűrűség- ill. eloszlásfüggvényt jelöli. Hasonló analitikus kifejezéseket kaphatunk más eloszlások esetén is. Ha ilyen zárt alak nem áll rendelkezésünkre, de tudunk F eloszlású véletlen számokat generálni, akkor Monte Carlo módszerek alkalmazhatók a jobb oldal kiszámítására.

A hozzátartozó várható szkór függvény (entrópia):

$$G(F) = - \int_{-\infty}^{\infty} F(y)(1 - F(y)) dy = -\frac{1}{2} E_F |X - X'| \quad (1.39)$$

míg a megfelelő divergencia Cramér-von Mises típusú:

$$d(F, G) = \int_{-\infty}^{\infty} (F(y) - G(y))^2 dy \quad (1.40)$$

Állítás ([2]): A *CRPS* lényeges a \mathcal{P} osztályon és szigorúan lényeges \mathcal{P} véges első momentummal rendelkező Borel-mértékekből álló \mathcal{P}_1 részosztályán.

Gyakran negatív orientációval használják, azaz $CRPS^*(F, x) = -CRPS(F, x)$ -vel számolnak és a kapott értékeket veszteségekként kezelik és minimalizálni szeretnék. Ekkor a $CRPS^* = E_F|X - x| - \frac{1}{2}E_F|X - X'|$ felírásból látszik, hogy az abszolút hiba általánosításával van dolgunk. Mikor F egy determinisztikus előrejelzés (x^*), azaz a megfelelő P egy pontmérték, akkor tényleg az abszolút hibát kapjuk vissza: $|x^* - x|$. Így e szkór használatával egy természetes módszer adódik determinisztikus és valószínűségi előrejelzések összehasonlítására. Azért hasznos ez az általánosítás, mert lényeges szkórt kaptunk, az abszolút hibát áttemelhetnénk máshogy is szkóros környezetbe, mint ahogy azt a következő részben bemutatom, de egyáltalán nem biztos, hogy lényeges szkórt fogunk kapni.

1.5. Az előrejelzői teljesítmény klasszikus mérőszámai

A következőkben az egyszerűség kedvéért tegyük fel, hogy az összes szóba kerülő momentum véges minden mérték esetében. Először legyen μ egy pontbecslés és legyen x a megfigyelt érték. Ekkor a két legegyszerűbb módszer a becslés hatékonyságának mérésére az *abszolút hiba*: $AE(\mu, x) = |\mu - x|$ és a *négyzetes hiba*: $SE(\mu, x) = (\mu - x)^2$. Ezek természetes módon általánosíthatók valószínűségi előrejelzések esetére is. Definiálhatjuk az *abszolút hiba szkórt* és a *négyzetes hiba szkórt* a következőképpen:

$$AES(P, x) = -|x - \mu_P|, \quad SES(P, x) = -(x - \mu_P)^2 \quad (1.41)$$

ahol μ_P a P prediktív eloszlás várható értéke.¹³ Az első nem lesz lényeges míg a második lényeges, de nem szigorúan lényeges, ugyanis a $E_P|X - x|$ kifejezést a P prediktív eloszlás mediánja minimalizálja, ezért például $\mu_Q = med_P$ esetben nem teljesül a lényegességet definiáló egyenlőtlenség:

$$AES(P, P) = -E_P|X - \mu_P| \leq -E_P|X - \mu_Q| = AES(Q, P) \quad (1.42)$$

Míg a $E_P(X - x)^2$ kifejezést μ_P minimalizálja, s így tetszőleges μ_Q mellett

$$SES(P, P) = -E_P(X - \mu_P)^2 \geq -E_P(X - \mu_Q)^2 = SES(Q, P) \quad (1.43)$$

Itt egyenlőség a $\mu_P = \mu_Q$ esetben áll, ezért csak lényeges és nem szigorúan lényeges szkórt kapunk.¹⁴

A SES azonos várható értékű prediktív eloszlásokat nem tud megkülönböztetni, így eléggé általános, az eloszlás laposságát, élességét, szórásnégyzetét ekkor még szabadon változtathatjuk. A következő *normalizált négyzetes hiba szkór* ezt hívatott orvosolni:

$$NSES(P, x) = \left(\frac{x - \mu_P}{\sigma_P} \right)^2 \quad (1.44)$$

ahol μ_P és σ_P a P prediktív eloszlás várható értéke és szórása.

¹³Mivel ezek negatívan orientáltak, egy negatív előjellel csinálunk belőlük pozitívan orientált szkórokat.

¹⁴Ennek alapján a $S(P, x) = |x - med_P|$ képlettel definiált szkór lesz lényeges, de nem szigorúan lényeges.

1.5.1. PMCC (Predictive Model Choice Criteria)

Képzeljük el, hogy a rendelkezésünkre álló x_1, \dots, x_n megfigyelések alapján valamilyen modellt illesztünk az adatokra és ennek a "minőségére" vagyunk kíváncsiak. A PMCC az illesztés hibáját a következőképp értelmezi:

$$PMCC = \sum_{i=0}^n (\mu_i - x_i)^2 + \sum_{i=1}^n \sigma_i^2 \quad (1.45)$$

ahol μ és σ^2 jelöli az illesztett modell várható értékét és szórásnégyzetét. Ezek általában nem függenek i -től, de előfordulhat, hogy minden x_i -hez tartozik egy megfigyelési súly, ekkor persze minden egyes mintaelemhez egy ezzel súlyozott prediktív eloszlás tartozik.

Ez az eddigi környezetben az

$$S(F, x) = -(E_F X - x)^2 - D_F^2(X) \quad (1.46)$$

szkórnak felel meg, és a véges második momentummal rendelkező F eloszlásfüggvényű eloszlásokon van értelmezve.

Könnnyen látható, hogy ez nem lesz lényeges, mivel a várható szkór nem akkor lesz maximális, ha a valódi F eloszlást adjuk előrejelzésnek, hanem az $E_F X$ pontbecslés esetében, ugyanis $S(E_F X, x) = -(E_F X - x)^2$ és így

$$S(E_F X, F) = -E_F (E_F X - X)^2 = -D_F^2(X) \geq -2D_F^2(X) = S(F, F) \quad (1.47)$$

Egyenlőség pedig csak $D_F^2(X) = 0$, azaz az $F = E_F X$ esetben lehetséges. Tehát nem a valódi háttéreloszlás esetén lesz minimális a várható szkór, hanem a háttéreloszlás várható értékének megfelelő pontbecslés esetén.

Az ebben a részben bemutatott szkórok a prediktív eloszlástól csak az első két momentumon keresztül függenek. Dawid és Sebastiani ([5], 1999) adott összefoglalást ezzel a tulajdonsággal rendelkező szkórokról. A következő *Dawid-Sebastiani szkór* például a PMCC egy lényeges alternatívájának tekinthető

$$DSS(P, x) = \left(\frac{x - \mu_P}{\sigma_P} \right)^2 + 2 \log \sigma_P \quad (1.48)$$

Normális prediktív eloszlás esetében ez ekvivalens a logaritmikus szkórral, ugyanis csak egy multiplikatív konstanstól eltekintve különböznek.

A CRPS alapján vizsgálhatjuk a

$$S(F, x) = - \int_0^1 (F^{-1}(u) - x)^2 du = -E_F (X - x)^2 \quad (1.49)$$

képlettel definiált szkórt is, viszont ez sem lesz lényeges.¹⁵

Első ránézésre úgy tűnhet, hogy csak egy $D_F^2(X)$ -es tagban különbözik a PMCC-től, ám itt a várható értéket a zárójelen kívül kell képezni, míg ott belül, így nem ekvivalensek. Az ilyen integrálalakban előállítható szkórokról bővebben majd a kvantiliselőjelzésekkel foglalkozó részben tesztek említést.

¹⁵A neki megfelelő fogalmak a Mallows-távolság és a Wasserstein-metrika.

1.6. Energia szkór

A következőben a CRPS egy igen széleskörű általánosítását fogom bemutatni, melyet az (1.38) analitikus előállítás tesz lehetővé. Jelölje \mathcal{P}_β , $\beta \in (0, 2)$ azon P valószínűségi Borel-mértékek családját \mathbb{R}^m -en, melyekre a $E_P\|X\|^\beta$ várható érték véges, ahol $\|\cdot\|$ jelöli az euklideszi normát. Ekkor az *energia szkórt* a következőképpen definiáljuk:

$$ES(P, x) = \frac{1}{2}E_P\|X - X'\|^\beta - E_P\|X - x\|^\beta \quad (1.50)$$

ahol X és X' független vektorváltozók a $P \in \mathcal{P}_\beta$ eloszlásból. Láthatóan a CRPS általánosítása, mely a $\beta = 1$ és $m = 1$ esetnek felel meg. Most megengedjük, hogy a változóink vektorértékűek legyenek és a β index is nagyobb tartományban mozoghat. A (1.51) kiszámítása diszkrét esetben egyszerű, különben meg Monte Carlo módszerekkel történhet. Negatív orientációval a β -rendű abszolút hiba $(|x^* - x|^\beta)$ általánosításának is tekinthető.

Állítás (Székely, [12], 2003): *Az energia szkór szigorúan lényeges \mathcal{P}_β -n.*

Az energia szkór az $\beta \in (0, 2)$ esetben az összes valószínűségi Borel-mértékre értelmezhető \mathbb{R}^m -en a következőképpen:

$$ES(P, x) = -\frac{\beta 2^{\beta-2} \Gamma(\frac{m+\beta}{2})}{\pi^{m/2} \Gamma(1 - \frac{\beta}{2})} \int_{\mathbb{R}^m} \frac{|\varphi(y) - e^{i\langle x, y \rangle}|^2}{\|y\|^{m+\beta}} dy \quad (1.51)$$

ahol φ jelöli a P mérték karakterisztikus függvényét. Így tehát a szkór a prediktív eloszlás és a megfigyelés karakterisztikus függvényének egy súlyozott távolságát számolja ki. Ha $P \in \mathcal{P}_\beta$, akkor (1.51) és (1.51) megegyezik (Székely, [12]). A $\beta = 2$ határesetben (1.51) jobboldala $E_P X$ és x négyzetes euklideszi távolságává redukálódik. Ekkor lényeges, de nem szigorúan lényeges szkórt kapunk a \mathcal{P}_2 eloszláscsaládon, melyekre $E_P\|X\|^2$ véges.

1.7. Lényeges szkórok és negatív definit függvények kapcsolata

Ebben a fejezetben negatív definit függvények segítségével definiálok lényeges szkórokat, megvizsgálom, hogy néhány eddig tárgyalt szkór, mely speciális esetnek felel meg, majd a CRPS egy igen messzemutató általánosítását mutatom be. Legyen Ω egy nemüres halmaz. Egy $\Omega \times \Omega$ -n értelmezett valós értékű g függvényt *negatív definit magnak* nevezünk, ha szimmetrikus és $\sum_{i=1}^n \sum_{j=1}^n a_i a_j g(x_i, x_j) \leq 0$ minden n pozitív egész, $x_1, \dots, x_n \in \Omega$, és olyan $a_1, \dots, a_n \in \mathbb{R}$ esetén, melyre $\sum_{i=1}^n a_i = 0$.

Tétel ([2]): *Legyen Ω egy Hausdorff tér és g egy nemnegatív, folytonos negatív definit mag $\Omega \times \Omega$ -n. Legyen P egy valószínűségi Borel-mérték Ω -n és legyen X és X' két független P eloszlású valószínűségi változó. Ekkor a*

$$S(P, x) = \frac{1}{2}E_P g(X, X') - E_P g(X, x) \quad (1.52)$$

képlettel definiált szkór lényeges a valószínűségi Borel-mértékek azon osztályán, melyekre az $E_P g(X, X')$ várható érték véges.

A jobb oldal kiszámítása ismét MCMC módszerrel történhet. Lássunk néhány példát, mind-egyikben a mintateret a standard topológiával látjuk el. Nézzük meg, hogy pár eddig látott szkór melyik speciális esetnek felel meg:

- Legyen $\Omega = \{0, 1\}$ és $g(0, 0) = g(1, 1) = 0$, $g(0, 1) = g(1, 0) = 2$. Ekkor visszkapjuk a Brier szkórt.
- Ha $\Omega = \mathbb{R}$ és $g(x, x') = |x - x'|$, akkor a CRPS-t kapjuk meg.
- Hasonlóan az $\Omega = \mathbb{R}^m$, $\beta \in (0, 2)$, $g(x, x') = \|x - x'\|^\beta$ esetben újra az energia szkórt kapjuk.
- Lássunk a CRPS egy eddig nem látott megfelelőjét (cirkuláris változókra):

Legyen $\Omega = \mathcal{S}^1$ a körvonal és jelölje $\alpha(\theta, \theta')$ két elem szögtávolságát. Legyen P egy valószínűségi Borel-mérték \mathcal{S}^1 -en és legyenek Θ és Θ' független P eloszlású változók. Mivel Gneiting ([2], 1998) egy tétele szerint a szögtávolság egy negatív definit mag $\mathcal{S}^1 \times \mathcal{S}^1$ -en, ezért a

$$S(P, \theta) = \frac{1}{2} E_P \alpha(\Theta, \Theta') - E_P \alpha(\Theta, \theta) \quad (1.53)$$

képlettel definiált szkór lényeges lesz a körvonal valószínűségi Borel-mértékeinek halmozán.

Most az energia szkór egy messzemutató általánosítását mutatom be, $x = (x_1, \dots, x_n) \in \mathbb{R}^m$ és $\alpha \in (0, \infty]$ esetén definiáljuk $\|x\|_\alpha$ -t a következőképpen:

$$\|x\|_\alpha = \begin{cases} (\sum_{i=1}^m |x_i|^\alpha)^{1/\alpha} & \alpha \in (0, \infty) \\ \max_{1 \leq i \leq m} |x_i| & \alpha = \infty \end{cases}$$

Schoenberg tétele ([2]) szerint ekkor $\alpha \in (0, \infty]$ és $\beta > 0$ esetén a következő $g(x, x') = \|x - x'\|_\alpha^\beta$ magfüggvény akkor és csak akkor negatív definit, ha

- (i) $m = 1, \alpha \in (0, \infty]$ és $\beta \in (0, 2]$ vagy
- (ii) $m \geq 2, \alpha \in (0, 2]$, és $\beta \in (0, \alpha]$ vagy
- (iii) $m = 2, \alpha \in (2, \infty]$, és $\beta \in (0, 1]$

feltételek egyike teljesül.

Ekkor tehát az előző tétel szerint a következő

$$S(P, x) = \frac{1}{2} E_P \|X - X'\|_\alpha^\beta - E_P \|X - x\|_\alpha^\beta \quad (1.54)$$

képlettel definiált szkór lényeges lesz \mathbb{R}^m valószínűségi Borel-mértékeinek véges $E_P \|X - X'\|_\alpha^\beta$ várható értékkel rendelkező részosztályán. Láthatóan az $\alpha = 2$ esetben visszkapjuk az energia szkórt, míg az $m \geq 2, \alpha \neq 2$ esetben nem-euklideszi analógokat kapunk.

Mattner ([2], 1997) megmutatta, hogy $\alpha \geq 1$ esetén $E_{P,Q} \|X - Y\|_\alpha^\beta$ akkor és csak akkor véges, ha $E_P \|X\|_\alpha^\beta$ és $E_Q \|Y\|_\alpha^\beta$ azok. Speciálisan $\alpha \geq 1$ esetén $E_P \|X - X'\|_\alpha^\beta$ akkor és csak akkor véges, ha $E_P \|X\|_\alpha^\beta$ az.

A következő tétel euklideszi mintatér esetén **szigorú** lényegességet biztosít bizonyos szkórokra, ehhez azonban előbb szükség van a teljesen monoton függvény fogalmára:

Definíció: Egy η $(0, \infty)$ -n értelmezett függvény *teljesen monoton*, ha mindegyik deriváltja létezik és ezekre $(-1)^k \eta^{(k)}(t) \geq 0$ teljesül minden k nemnegatív egészre és $t > 0$ -ra.

Tétel ([2]): Legyen Ψ folytonos függvény $[0, \infty)$ -n, melyre $-\Psi'$ teljesen monoton és nem konstans. Legyen P egy valószínűségi Borel-mérték \mathbb{R}^m -en és legyen X és X' két független P eloszlású vektorváltozó. Ekkor a

$$S(P, x) = \frac{1}{2} E_P \Psi(\|X - X'\|_2^2) - E_P \Psi(\|X - x\|_2^2) \quad (1.55)$$

képlettel definiált szkór **szigorúan lényeges** lesz a valószínűségi Borel-mértékek azon részosztályán, melyre a $E_P \Psi(\|X - X'\|_2^2)$ várható érték véges.

Speciálisan a $\Psi(t) = t^{\beta/2}$ ($\beta \in (0, 2)$) választással az energia szkór szigorú lényegességét kaphatjuk meg a véges $E_P \|X\|_2^\beta$ várható értékkel rendelkező valószínűségi Borel-mértékek osztályán.

1.8. Szkórok intervallum- és kvantiliselőrejelzések esetében

Nem minden esetben van szükségünk egy egész eloszlásra, lehet hogy elég lenne nekünk csak egy 95 százalékos kvantilis, például hipotézisvizsgálathoz. S mivel néha igen nehéz egy egész eloszlást meghatározni, egy előrejelzés történhet prediktív intervallumokon illetve prediktív kvantiliseken keresztül is.¹⁶ Most ezt a két esetet részletezem:

1.8.1. Lényeges szkórok kvantilis előrejelzésekhez

Most egy egész eloszlás helyett csak néhány kvantilist adunk meg előrejelzésként. Ezeket *prediktív kvantiliseknek* nevezzük. Vegyünk $\alpha_1, \dots, \alpha_k \in (0, 1)$ elemeket, az ilyen szintű kvantilisek előrejelzésére törekszünk. Most egy szkór a prediktív kvantiliseken és a megfigyelésen van értelmezve, azaz ha egy előrejelző az r_1, \dots, r_k előrejelzést bocsátja ki és x a megfigyelés, akkor az ő jutalma $S(r_1, \dots, r_k; x)$ lesz. A *várható szkórt* a P valószínűségi mérték és az r_1, \dots, r_k előrejelzések mellett a

$$S(r_1, \dots, r_k; P) = \int S(r_1, \dots, r_k; x) dP(x) \quad (1.56)$$

képlettel definiáljuk. A technikai nehézségek elkerülése végett tegyük fel, hogy $P \in \mathcal{P}$, ahol \mathcal{P} olyan Borel-mértékek konvex osztálya \mathbb{R} -en, melyeknek minden momentuma véges és eloszlásfüggvényük szigorúan monoton.

Egy $P \in \mathcal{P}$ esetén jelölje q_1, \dots, q_k az $\alpha_1, \dots, \alpha_k \in (0, 1)$ szinteknek megfelelő valódi P -kvantiliseket. Ekkor az S szkórt *lényegesnek* nevezzük, ha $\forall r_1, \dots, r_k \in \mathbb{R}$ és $\forall P \in \mathcal{P}$ esetén

$$S(q_1, \dots, q_k; P) \geq S(r_1, \dots, r_k; P) \quad (1.57)$$

¹⁶Például Krzysztofowicz és Sigrest ([11], 1999) csapadékmennyiség kvantiliselőrejelzéseinek foglalkozott.

A következőkben megadom a lényeges szkórok egy igen széles osztályát, melyek kvantiliselőrejelzések kiértékelésére alkalmasak. Először egyetlen kvantilis előrejelzése esetén majd tetszőleges számúra általánosan. Tegyük fel, hogy fellépő $s(x)$ és $h(x)$ függvények \mathcal{P} -mérhetőek és legfeljebb polinomiálisan nőnek x -ben. Ekkor érvényes a következő:

Tétel ([2]): *Ha $s(x)$ monoton növekszik és $h(x)$ tetszőleges, akkor az*

$$S(r; x) = \alpha s(r) + (s(x) - s(r)) \mathbf{1}\{x \leq r\} + h(x) \quad (1.58)$$

képlettel definiált szkór lényeges az α -szintű kvantilis előrejelzése esetében.

Bizonyítás: Legyen q az (egyetlen) α -kvantilise a $P \in \mathcal{P}$ valószínűségi mértéknek, ekkor a következő jelölést használjuk: $P(q) = \alpha$. Ha $r < q$, akkor

$$\begin{aligned} S(q; P) - S(r; P) &= \int_{(r, q)} s(x) dP(x) + s(r)P(r) - \alpha s(r) \\ &\geq s(r)(P(q) - P(r)) + s(r)P(r) - \alpha s(r) = 0 \end{aligned}$$

ahogy állítottuk. Az $r > q$ eset hasonlóan intézhető el. \square

A leggyakoribb eset, mikor $s(x) = x$ -t és $h(x) = -\alpha x$ -t helyettesítenek a képletbe. Ekkor a következő lényeges szkórt kapjuk¹⁷:

$$S(r; x) = (x - r)(\mathbf{1}\{x \geq r\} - \alpha) = \begin{cases} (1 - \alpha)(x - r) & \text{ha } x \geq r \\ \alpha(r - x) & \text{ha } x < r \end{cases} \quad (1.59)$$

Következmény *Ha s_i ($i = 1, \dots, k$) monoton növekvő függvények, és h tetszőleges, ekkor az*

$$S(r_1, \dots, r_k; x) = \sum_{i=1}^k (\alpha_i s_i(r_i) + (s_i(x) - s_i(r_i)) \mathbf{1}\{x \leq r_i\}) + h(x) \quad (1.60)$$

szkór lényeges az α_i -szintű kvantilisek előrejelzése esetén. ($i = 1, \dots, k$)

Cervera és Muñoz ([8], 1996) még csak lineáris s_i -k esetében bizonyították ezt, és felvetették a kérdést, hogy vajon csak ezek-e lényegesek kvantiliselőrejelzések esetén. Gneiting és Raftery fenti következménye adta meg a negatív választ erre, de ők sem adtak végleges választ arra, hogy vajon a fenti képletek szolgáltatják-e a lényeges szkórok általános alakját.

Mivel egy eloszlásfüggvény megadása ekvivalens az összes kvantilis megadásával, így könnyen kaphatunk különböző szkórokot eloszlásfüggvény-előrejelzések esetére a megfelelő kvantilis-szkórokból. Ehhez legyen ν tetszőleges Borel-mérték a $(0, 1)$ intervallumon, és S_α jelöljön egy az α szinthez tartozó lényeges kvantilis-szkórt, ekkor a következő

$$S(F, x) = \int_0^1 S_\alpha(F^{-1}(\alpha); x) \nu(d\alpha) \quad (1.61)$$

eloszlásfüggvények előrejelzésének kiértékelésére használt szkór szintén lényeges megfelelő integrálhatósági feltételek mellett.

¹⁷Ekkor tehát épp az α szintnek megfelelő felosztásnak megfelelően büntetjük, hogyha alul- vagy felülbecsüljük a kvantilist.

Hasonlóan építhetünk fel szkórokat prediktív eloszlásfüggvényekhez bináris előrejelzésekhez tartozó szkórokból. Legyen S egy bináris előrejelzésekhez tartozó lényeges szkór, és ν tetszőleges Borel-mérték a számegeyenesen. Ekkor az

$$S^*(F, x) = \int_{-\infty}^{\infty} S(F(y), \mathbf{1}\{y \geq x\}) \nu(dy) \quad (1.62)$$

formulával definiált szkór is lényeges, ha az integrál létezik és véges. Emlékezzünk vissza, hogy a CRPS-t is így kaptuk a megfelelő bináris Brier szkórból, a megfelelő ν a Lebesgue-mérték volt. Ez a konstrukció működik a többdimenziós esetben is: legyen \mathcal{P} Borel-mértékek egy osztálya \mathbb{R}^m -en, és egy előrejelzést adjunk meg az \mathcal{O} m -dimenziós eloszlásfüggvényével. Ekkor a CRPS többváltozós analógja a következőképpen értelmezhető:

$$CRPS(F, x) = - \int_{\mathbb{R}^m} (F(y) - \mathbf{1}\{y \geq x\})^2 \nu(dy) \quad (1.63)$$

Ez a kvadratikus szkór egy súlyozott integrálja mind az m változóban, így a ν Borel-mérték választható olyan szorzatmértéknek, hogy az előrejelző a fontosabbakra koncentrálja a figyelmét. Ha ν véges, és a Lebesgue-mérték dominálja, akkor ez szigorúan lényeges a \mathcal{P} osztályon.

1.8.2. Intervallumszkór

Az intervallumelőrejelzések egy igen fontos speciális esetét képezik a kvantiliselőrejelzéseknek. A klasszikus centrális $(1 - \alpha) \times 100$ % prediktív intervallumok esetével foglalkozom, melyek alsó határpontja az $\frac{\alpha}{2}$ szintű kvantilis egy előrejelzése, míg a felső az $(1 - \frac{\alpha}{2})$ szintű prediktív kvantilis. Ekkor jelöljük egy S szkórt a következőképpen: $S_\alpha(l, u; x)$, ahol l és u jelöli a megfelelő kvantiliselőrejelzéseket, x pedig a megfigyelést. Vagyis ha egy előrejelző az $[l, u]$ intervallumelőrejelzést adja, és x -et figyeljük meg, akkor az \mathcal{O} jutalma $S_\alpha(l, u; x)$. Ekkor az előző tétel jelöléseit használva, az $\alpha_1 = \frac{\alpha}{2}$, $\alpha_2 = 1 - \frac{\alpha}{2}$ és az $s_1(x) = s_2(x) = 2\frac{x}{\alpha}$ és a $h(x) = -\frac{x}{\alpha}$ választás következő intervallumszkórt eredményezi:

$$S_\alpha^{int}(l, u; x) = - \left((u - l) + \frac{2}{\alpha}(l - x)\mathbf{1}\{x < l\} + \frac{2}{\alpha}(x - u)\mathbf{1}\{x > u\} \right) \quad (1.64)$$

Ez az intervallumszkór szűk prediktív intervallumok előrejelzésére bízgat, és elkerülünk egy az α -tól függő büntetést, ha az intervallum lefedi a megfigyelést. Hasonlóan kaphatunk más lényeges szkórokat is az $s_1(x)$, $s_2(x)$ és $h(x)$ különböző megválasztásával.

Az intervallumszkór a gyakorlatban [7]

Vegyük a következő X_t stacionárius bilineáris folyamatot, ahol ϵ_t független $\mathcal{N}(0, 1)$ zaj

$$X_{t+1} = \frac{1}{2}X_t + \frac{1}{2}X_t\epsilon_t + \epsilon_t \quad (1.65)$$

Kabaila és He [7] egy lépéses centrális intervallumelőrejelzéseket vizsgált 95%-os szinten. A folyamat Markov jellegű, így X_{t+1} eloszlása adott X_t , X_{t-1}, \dots mellett normális $\frac{1}{2}X_t$ várható értékkel és $(1 + \frac{1}{2}X_t)^2$ szórásnégyzettel, melyből a következő intervallumbecslés kapható:

$$I = \left[\frac{1}{2}X_t - c \left| 1 + \frac{1}{2}X_t \right|, \frac{1}{2}X_t + c \left| 1 + \frac{1}{2}X_t \right| \right] \quad (1.66)$$

ahol $c = \Phi^{-1}(0.975)$. Ezt az I -t elnevezték standard prediktív intervallumnak, mert az alsó és felső határpontja a feltételes eloszlás 2.5%-os ill. 97.5%-os kvantilise.

Emellett még a következő két másik alternatívát vizsgálták:

$$J = [F^{-1}(0.025), F^{-1}(0.975)] \quad (1.67)$$

ahol F jelöli a folyamat feltétel nélküli (stacionárius) eloszlásának eloszlásfüggvényét, és

$$K = \left[\frac{1}{2}X_t - \gamma \left(\left| 1 + \frac{1}{2}X_t \right| \right), \frac{1}{2}X_t + \gamma \left(\left| 1 + \frac{1}{2}X_t \right| \right) \right] \quad (1.68)$$

ahol

$$\gamma(y) = \begin{cases} \left(2 \log \left(\frac{7.36}{y} \right) \right)^{1/2} y & \text{ha } 0 < y \leq 7.36 \\ 0 & \text{ha } y > 7.36 \end{cases}$$

Ez kicsit félrevezetőnek tűnhet, mert a feltételes prediktív szórásnégyzet nagy értéke mellett egy pontbecsléssé redukálódik, ám γ ezen választása adott nominális fedettség mellett minimalizálja a prediktív intervallum várható szélességét. Egy 100.000 elemszámú mintát generálva a bilineáris folyamatból az intervallumelőrejelzésekre a következő eredményt kapták:

Intervallum- előrejelzés	Tapasztalati lefedettség	Átlagos szélesség	Intervallum- átlagszór
I	95.01%	4.00	-9.55
J	95.08%	5.45	-16.09
K	94.98%	3.79	-10.64

Mindhárom esetben a tapasztalati lefedettség igen közel mutatkozik a nominálishoz, a második K intervallumbecslésnek legkisebb az átlagos szélessége, de végülis az I standard prediktív intervallum adja a legjobb eredményt az intervallumszór használatakor.

1.9. Optimális szór becslés

Mint már említettem, a lényeges szókörök fontos szerephez jutnak a becslélméletben is. Képzeljük el, hogy az X_1, \dots, X_n független, azonos eloszlású megfigyelések alapján akarunk egy P_ϑ paraméteres modellt illeszteni az adatokra valamilyen $\{P_\vartheta : \vartheta \in \Theta\}$ családból. Mikor ϑ -t becsljük, mérhetjük a becslés jószágát az átlagos szkórral:

$$\mathcal{S}_n(\vartheta) = \frac{1}{n} \sum_{i=1}^n S(P_\vartheta, X_i) \quad (1.69)$$

ahol S egy (szigorúan) lényeges szór valószínűségi mértékek egy konvex osztályán, mely tartalmazza a paraméteres modellt. Ha ϑ_0 jelöli a valódi paraméterértéket, akkor enyhe feltételek mellett bebizonyítható, hogy $\arg \max_{\vartheta} \mathcal{S}_n(\vartheta) \rightarrow \vartheta_0$. Így egy természetes becslési eljárást kapunk: válasszunk ki egy a problémához illő S szigorúan lényeges szkórt, majd maximalizáljuk $\mathcal{S}_n(\vartheta)$ -t a Θ paraméterterén. Az így kapott $\hat{\vartheta}_n = \arg \max_{\vartheta} \mathcal{S}_n(\vartheta)$ konzisztens becslést az \mathcal{S} -hez tartozó *optimális szór becslésnek* nevezzük. Ez a becslésoosztály nagyon sok idáig is ismert becsléseket foglal magába, mint például a maximum likelihood-becslés, legkisebb négyzetes becslés vagy más regressziós becslések. Valahol a maximum likelihood-becslés és az M -becslések között helyezkednek el félúton: a ML-becslés a logaritmikus szkórnak felel meg, míg az M -becslésnél optimalizálandó függvény esetünkben egy szigorúan lényeges szkórból ered. Így az M -becslésekre bizonyított tulajdonságok itt is érvényesek, például az

aszimptotikus viselkedés is öröklődik. Például vegyük a normális eloszlás eltolásparaméterének a CRPS-ből eredő optimális szkór becslését ismert szórás esetén. Ekkor az M -becsléshez tartozó ψ függvény a következő alakú $\psi(x) = 2\Phi(\frac{x}{c}) - 1$ ahol c pozitív konstans.

Nagy előnyük, hogy a feladat természetétől függően mi választhatjuk meg melyik szkórhoz tartozó hasznossági függvénnyel dolgozunk. Pfanzagl, Birgé és Massart [2] *optimális kontraszt becslés* néven foglalkozott velük. Bebizonyították a konzisztencia egy szigorúbb verzióját, és megmutatták, hogy a konvergencia sebessége a paraméterter entrópikus struktúrájával van kapcsolatban. Számos probléma esetében kiderült, hogy a ML-becslés nem elég jó, ezért érdemes a logaritmikusság helyett valamilyen a problémához jobban illő S szkórból eredő optimális szkór becsléssel próbálkozni.

A pontbecslések mellett persze vizsgálhatunk intervallumbecsléseket is az intervallumszkórból kiindulva. A következő paradoxonra J. O. Berger [2] hívta fel a figyelmet. Egy normális populáció eltolásparaméterére keresünk intervallumbecsléseket, ismeretlen skálaparaméter mellett. Legyen

$$L(I, \theta) = c\lambda(I) - \mathbf{1}\{\theta \in I\} \quad (1.70)$$

a veszteségfüggvényünk, ahol c pozitív konstans és $\lambda(I)$ jelöli az I intervallumbecslés Lebesgue-mértékét. Első ránézésre nagyon intuitívnak tűnhet L használata, hiszen minél keskenyebb intervallumbecslésekre biztat, de bünteti, hogyha az intervallum nem tartalmazza a paramétert. A baj, hogy ezzel a klasszikus t-próbának megfelelő konfidenciaintervallumot dominálja egy félrevezető intervallumbecslés, mely nagyon rossz tulajdonságokkal rendelkezik, például nagyon gyorsan ráhúzódik a mintaátlagra a legbizonytalanabb helyzetekben is. Pedig az L figyelembe veszi az intervallumunk szélességét és lefedettséget is. Mégis azt a következtetést vonhatjuk le, hogy az egyetlen kiváltó oka ennek a nemkívánatos viselkedésnek csak az lehet, hogy valami baj van a veszteségfüggvénnyel. Ezért nagyon ajánlott itt is a lényeges szkórokat elővenni és belőlük származó veszteségfüggvényekkel dolgozni. Először is két intervallumbecslés összehasonlítása igazán akkor bír értékkel, ha legalább az egyik a szélesség és a lefedettség közül le van rögzítve. Az L veszteségfüggvény minden halmazértékű becslésre alkalmazható függetlenül azok lefedettségétől, ill. mértéküktől, ami elég szükségtelennek, túl általánosnak tűnik. Ezért nézzünk olyan intervallumbecsléseket melyeknek a nominális lefedettsége megegyezik, azaz nézzük a legalább $(1 - \alpha)$ -megbízhatósági szintű szimmetrikus konfidenciaintervallumokat és használjuk az intervallumszkórnak megfelelő veszteségfüggvényt:

$$L_\alpha(I, \theta) = \lambda(I) + \frac{2}{\alpha} \inf_{\eta \in I} |\theta - \eta| \quad (1.71)$$

Észrevehetjük hogy az L_α veszteségfüggvény mellett sokkal rugalmasabban büntetünk, ugyanis figyelembe vesszük az intervallumbecslésünk és a becsülni kívánt paraméter távolságát.

2. fejezet

Kalibráció és élesség

2.1. Kalibráció

Gneiting, Raftery és mások a valószínűségi előrejelzések célját a következőképpen fogalmazzák meg: maximalizálni az eloszlás élességét rögzített kalibráció mellett. A kalibráció az adatok konzisztenciáját jelenti, mennyire egyeztethető össze statisztikailag a prediktív eloszlásunk a megfigyelésekkel, így közös tulajdonsága az előrejelzésnek és a megfigyeléseknek. Míg az élesség a prediktív eloszlás koncentrátságát fejezi ki, így csak magának az eloszlásnak a tulajdonsága. Minél koncentráltabb a prediktív eloszlás, annál élesebb az előrejelzés; és minél élesebb egy előrejelzés, annál jobb, adott kalibráció mellett. Először a valószínűségi előrejelzések kalibrációjával foglalkozom. A kalibráció különféle változatait definiálom, melyekről megmutatom, hogy logikailag függetlenek; majd különböző eszközöket mutatok be ezek meglétének ellenőrzésére illetve becslésére, mint például a marginális kalibráció diagramm, boxplot vagy a PIT (Probability Integral Transform), utóbbinak definiálom két, diszkrét előrejelzések esetére való általánosítását is (randomizált és nemrandomizált PIT). Végül szükséges és elégséges feltételt adok két megismert kalibráció meglétére egy F_t prediktív eloszlásfüggvény-sorozat esetén.

Ezután rátérek a prediktív eloszlások koncentrátságának a vizsgálatára, két példát mutatok arra, hogy a gyenge kalibráció mellett egy előrejelző lehet élesebb, mint az ideális előrejelző. Klimatológiai előrejelzés esetében azonban alsó korlátot biztosítható a prediktív eloszlás varianciájára.

Mindezeket példákon keresztül mutatom be, különböző ábrákkal illusztrálom a megismert fogalmakat, és a megismert lényeges szövegeket is felhasználom különböző előrejelzők rangsorolására.

Legyenek $(F_t)_{t=1,2,\dots}$ és $(G_t)_{t=1,2,\dots}$ folytonos és szigorúan monoton eloszlásfüggvények sorozatai, melyek függhetnek véletlen paramétereiktől is. $(G_t)_{t=1,2,\dots}$ -re úgy gondolunk, mint a valódi, természet által generált háttér folyamatra, míg $(F_t)_{t=1,2,\dots}$ -re mint valószínűségi előrejelzések sorozatára.

A következő definíció a két sorozat közt esetlegesen meglévő aszimptotikus kompatibilitásra utal:¹

¹Ha véletlen paramétereink is vannak, akkor mindenhol a konvergencia 1 valószínűséggel értendő.

Definíció (A kalibráció három fő fajtája)

(i) Az $(F_t)_{t=1,2,\dots}$ sorozat *valószínűségi kalibrált* a $(G_t)_{t=1,2,\dots}$ sorozathoz, ha

$$\frac{1}{T} \sum_{t=1}^T G_t \circ F_t^{-1}(p) \rightarrow p \quad \forall p \in (0, 1) \quad (2.1)$$

(ii) Az $(F_t)_{t=1,2,\dots}$ sorozat *exceedance kalibrált*² a $(G_t)_{t=1,2,\dots}$ sorozathoz, ha

$$\frac{1}{T} \sum_{t=1}^T G_t^{-1} \circ F_t(x) \rightarrow x \quad \forall x \in \mathbb{R} \quad (2.2)$$

(iii) Az $(F_t)_{t=1,2,\dots}$ sorozat *marginálisan kalibrált* a $(G_t)_{t=1,2,\dots}$ sorozathoz, ha a következő két határérték

$$\bar{G}(x) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T G_t(x) \right) \quad \text{és} \quad \bar{F}(x) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T F_t(x) \right) \quad (2.3)$$

létezik és egyenlő $\forall x \in \mathbb{R}$ és eloszlást határoznak meg.³

(iv) Az $(F_t)_{t=1,2,\dots}$ sorozat *erősen kalibrált* a $(G_t)_{t=1,2,\dots}$ sorozathoz, ha mind valószínűségi, exceedance és marginálisan kalibrált hozzá.

Ha az $(F_t)_{t=1,2,\dots}$ sorozat minden részsorozata valószínűségi kalibrált a $(G_t)_{t=1,2,\dots}$ megfelelő részsorozatához, akkor azt mondjuk hogy $(F_t)_{t=1,2,\dots}$ *teljesen valószínűségi kalibrált* $(G_t)_{t=1,2,\dots}$ -hez. Hasonlóan beszélhetünk *teljes exceedance*, *teljes marginális* és *teljes erős kalibráltságról* is.

A következőkben bemutatok néhány tipikus előrejelzőt és megvizsgálom őket kalibráltság szempontjából. Ehhez legyenek $(\mu_t)_{t=1,2,\dots}$, $(\sigma_t)_{t=1,2,\dots}$ és $(\tau_t)_{t=1,2,\dots}$ egymástól is független, független, azonos eloszlású valószínűségi változók sorozatai.

Mindegyik példában a G_t az $\mathcal{N}(\mu_t, 1)$ eloszlásfüggvénye lesz, ahol $\mu_t \sim \mathcal{N}(0, 1)$. Azaz minden pillanatban a természet kisorsol egy $\mathcal{N}(0, 1)$ eloszlású véletlen számot és a $\mathcal{N}(\mu_t, 1)$ eloszlást veszi háttéreloszlásul.⁴

Ideális előrejelző

Az ideális előrejelző prediktív eloszlása minden egyes időpontban megegyezik a természet eloszlásával, azaz $F_t = G_t$ minden t -re. Az ideális előrejelző természetesen erősen kalibrált.

Klimatológiai előrejelző

A klimatológiai előrejelző az $F_t = \mathcal{N}(0, 2)$ prediktív eloszlást bocsátja ki t -től függetlenül. Ez az előrejelző valószínűségi és marginálisan kalibrált, de nem exceedance, ugyanis

$$\frac{1}{T} \sum_{t=1}^T G_t^{-1} \circ F_t(x) = \frac{1}{T} \sum_{t=1}^T \left[\Phi^{-1} \left\{ \Phi \left(\frac{x}{\sqrt{2}} \right) \right\} + \mu_t \right] \rightarrow \frac{x}{\sqrt{2}} \neq x \quad (2.4)$$

²Az angol elnevezést használom, mivel nem találtam neki igazán megfelelő magyar kifejezést. A másik két kifejezés angol megfelelője *probabilistic calibration* és *marginal calibration*.

³A \bar{G} létezése alapvető feltétel meteorológiai problémákban és a stabil klíma létezéséhez kapcsolható. Lényegében a marginális kalibráció a megfigyelt és az előrejelzett klíma egyenlőségét hivatott kifejezni.

⁴ G_t eloszlása kevert normális: $G_t = G|\mu_t \sim \mathcal{N}(\mu_t, 1)$. Ekkor G feltétel nélküli eloszlása $\mathcal{N}(0, 2)$, ugyanis $f_G(x) = \int f_{G|\mu_t=y}(x|y) \cdot f_{\mu_t}(y) dy = f_{\mathcal{N}(0,2)}$

Vegyük észre, hogy az F_t prediktív eloszlás minden esetben megegyezik a \bar{G} feltétel nélküli eloszlással. Minden ezzel a tulajdonsággal rendelkező előrejelzőt klimatológiai előrejelzőnek nevezünk. Klimatológiai előrejelzés esetében a valószínűségi kalibráltság ekvivalens a marginális kalibráltsággal. Ugyanis, ha G folytonos és szigorúan monoton, akkor $p = F_t(x) = \bar{G}(x)$ -t helyettesítve (2.1)-be, a marginális kalibráltság definícióját kapjuk, ezt visszafele alkalmazva pedig a valószínűségi kalibráltságot. A gyakorlatban a klimatológiai előrejelzések történelmi megfigyelések alapján készülnek és gyakran mint referencia-előrejelzések használatosak.

Nem-fókuszált előrejelző

A nem-fókuszált előrejelző prediktív eloszlása a következő keverék eloszlás:

$$F_t \sim \frac{1}{2}\{\mathcal{N}(\mu_t, 1) + \mathcal{N}(\mu_t + \tau_t, 1)\} \quad (2.5)$$

ahol $\tau_t \frac{1}{2} - \frac{1}{2}$ valószínűséggel a $+1$ és -1 értékeket veszi fel, μ_t -től függetlenül.

Ekkor $F_t(x) = \frac{1}{2}\{\Phi_+(x - \mu_t) + \Phi_-(x - \mu_t)\}$, ahol $\Phi_{\pm}(x) = \frac{1}{2}\{\Phi(x) + \Phi(x \mp 1)\}$.

Vizsgáljuk meg ezt az előrejelzőt kalibráltság szempontjából, először nézzük a valószínűségi kalibráció (2.1) feltételét:

$$\frac{1}{T} \sum_{t=1}^T G_t \circ F_t^{-1}(p) \rightarrow \frac{1}{2}\{\Phi \circ \Phi_+^{-1}(p) + \Phi \circ \Phi_-^{-1}(p)\} = p \quad (2.6)$$

ahol az utolsó egyenlőség $p = \Phi_+(x)$ helyettesítéssel és egyszerűsítéssel adódik.⁵

Az exceedance kalibrációhoz (2.2) nem teljesül általában:

$$\frac{1}{T} \sum_{t=1}^T G_t^{-1} \circ F_t(x) \rightarrow \frac{1}{2}\{\Phi^{-1} \circ \Phi_+(x) + \Phi^{-1} \circ \Phi_-(x)\} \neq x \quad (2.7)$$

A marginális kalibráció (2.3) feltétele sem teljesül, ugyanis

$$\bar{G}(x) \sim \mathcal{N}(0, 2) \neq \bar{F} \sim \frac{1}{2}\mathcal{N}(0, 2) + \frac{1}{4}\mathcal{N}(-1, 2) + \frac{1}{4}\mathcal{N}(1, 2)$$

Tehát ez az előrejelző valószínűségi kalibrált, viszont se nem exceedance, se nem marginálisan kalibrált.

Átlag-torzított előrejelző

Ez az előrejelző a következő torzított előrejelzést bocsátja ki: $F_t \sim \mathcal{N}(\mu_t + \tau_t, 1)$, ahol τ_t ugyanaz mint az előző példában. Ez az előrejelző exceedance kalibrált, de se nem valószínűségi, se nem marginális kalibrált.

Előjel-torzított előrejelző

Ennek az előrejelzőnek a prediktív eloszlása: $F_t \sim \mathcal{N}(-\mu_t, 1)$. Ez az előrejelző exceedance és marginálisan kalibrált, viszont nem valószínűségi kalibrált.

Kevert előrejelző

A kevert előrejelző minden t időpontban a klimatológiai és az előjel-torzított előrejelzés közül választ $\frac{1}{2} - \frac{1}{2}$ valószínűséggel, μ_t -től függetlenül. Ez az előrejelző valószínűségi és exceedance

⁵Ugyanis $\Phi \circ \Phi_-^{-1} \circ \Phi_+(x) = \Phi(x - 1)$ és ezt átrendezve $\Phi_+(x) = \Phi_-(x - 1)$, amiből már következik az egyenlőség.

Kalibráció	Előrejelző
VEM	ideális előrejelző
VEM	$G_t = F_t = \mathcal{N}(t, 1)$
$V\bar{E}M$	klimatológiai előrejelző
$\bar{V}EM$	átlag-torzított előrejelző
$\bar{V}EM$	előjel-torzított előrejelző
$\bar{V}EM$	kevert előrejelző
$V\bar{E}M$	nem-fókuszált előrejelző
$\bar{V}EM$	Hamill-előrejelző

2.1. táblázat. A 3 kalibrációfogalom logikailag független

kalibrált, de nem marginálisan kalibrált.

Hamill-előrejelző

Ehhez az előrejelzőhöz a következő kevert prediktív eloszlás tartozik: $\mathcal{N}(\mu_t + \delta_t, \sigma_t^2)$, ahol $(\delta_t, \sigma_t^2) = (\frac{1}{2}, 1)$, $(-\frac{1}{2}, 1)$ vagy $(0, \frac{169}{100})$ egyforma valószínűséggel.⁶

$$\frac{1}{T} \sum_{t=1}^T G_t \circ F_t^{-1}(p) \rightarrow \frac{1}{3} \left[\Phi \left\{ \Phi^{-1}(p) - \frac{1}{2} \right\} + \Phi \left\{ \frac{13}{10} \Phi^{-1}(p) \right\} + \Phi \left\{ \Phi^{-1}(p) + \frac{1}{2} \right\} \right] = p + \varepsilon(p) \quad (2.8)$$

ahol $|\varepsilon(p)| \leq 0.0032$ minden p -re, de $\varepsilon(p) \neq 0$ általában. A valószínűségi kalibráció feltétele tehát megsérült, emellett az exceedance kalibráció (2.2) feltétele sem teljesül:

$$\frac{1}{T} \sum_{t=1}^T G_t^{-1} \circ F_t(x) \rightarrow \frac{1}{3} \left\{ \left(x + \frac{1}{2} \right) + \frac{10}{13}x + \left(x - \frac{1}{2} \right) \right\} = \frac{12}{13}x \quad (2.9)$$

Sőt ez az előrejelző marginálisan sem lesz kalibrált, ugyanis a két marginális eloszlás:

$$\bar{G} \sim \mathcal{N}(0, 2), \text{ míg } \bar{F} \sim \frac{1}{3} \left\{ \mathcal{N} \left(-\frac{1}{2}, 2 \right) + \mathcal{N} \left(\frac{1}{2}, 2 \right) + \mathcal{N} \left(0, \frac{269}{100} \right) \right\}$$

A fenti példák mutatják, hogy a három bevezetett kalibrációfogalom logikailag független egymástól, azaz bármilyen kombinációban feltűnhetnek.

2.1.1. Valószínűségi kalibráció becslése

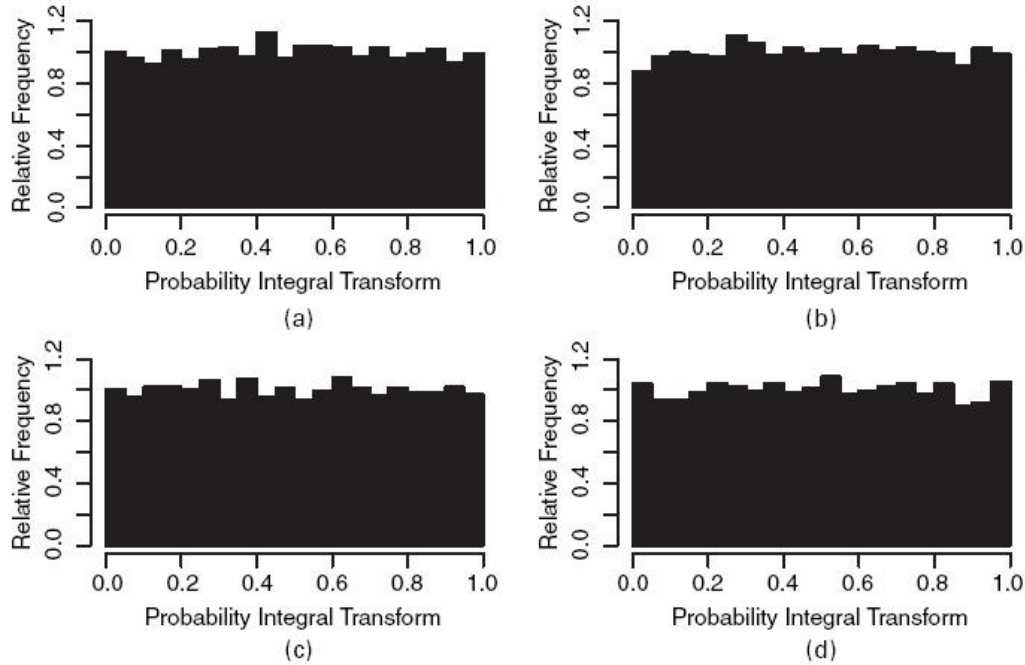
PIT (Probability Integral Transform)

A következő transzformáció azon alapszik, hogy egy x G folytonos eloszlású megfigyelést visszahelyettesítünk az F_G eloszlásfüggvényébe, akkor a $(0, 1)$ -n egyenletes eloszlású véletlen számot kapunk. Így ha képezzük a $p_t = F_t(x_t)$ úgynevezett *PIT-értékeket*, akkor ezek egyenletesen oszlanak el az ideális előrejelző esetében. Tehát ezek egyenletessége szükséges feltétele az ideális előrejelzésnek. Ha ez fordítva is igaz lenne, akkor mérhetnénk egy előrejelzésünk jóságát a PIT-értékek egyenletességével, de sajnos ez nincs így, olyannyira, hogy Hamill (2001) példát adott olyan előrejelzőre, mely PIT-értékeinek hisztogramja lényegében egyenletes, habár minden egyes előrejelzés torzított. Ebből kifolyólag a PIT-értékek

⁶Itt a (δ_t, σ_t^2) μ_t -től és egymástól is független sorozatot alkot.

egyenletessége csak szükséges, de nem elégséges feltétele az ideális előrejelzőnek.⁷

A következő ábrán egy 10.000 elemszámú mintából generált PIT-hisztogram látható 4 megismert előrejelző esetén, mindegyik esetben a $[0, 1]$ intervallumot beosztjuk 20 egyenlő részre, majd ábrázoljuk ezek relatív gyakoriságát. Láthatjuk, hogy nagyjából mindegyik egyenletes:



2.1. ábra. (a) ideális, (b) klimatológiai, (c) nem-fókuszált, (d) Hamill előrejelző

A PIT-hisztogram vizuális vizsgálata magyarázatot adhat az előrejelzés hiányosságaira. Domború hisztogram túlságosan "széles" prediktív eloszlásra utal, túl széles prediktív intervallumokkal. Homorú hisztogram esetében viszont túlságosan is "szűk" prediktív eloszlásra kell gyanakodnunk. Háromszög-alakú hisztogramot pedig akkor láthatunk, ha a prediktív eloszlásunk túl ferde.

Előrejelző	Nominális lefedettség	
	50%	90%
Ideális	51.2	90.1
Klimatológiai	51.3	90.7
Nem-fókuszált	50.4	90.3
Hamill	50.8	89.7

2.2. táblázat. A prediktív intervallumok empirikus lefedettsége

A (2.2) táblázat összefoglalja a megfelelő 50% és 90% prediktív intervallumok empirikus lefedettségét. Ez a PIT-hisztogram középső 10 illetve 18 beosztás alatti részéből is leolvasható.

⁷ Az irodalom gyakran Rosenblatt (1952) nevéhez fűzi, de egészen Pearson (1933) munkásságáig visszavezethető.

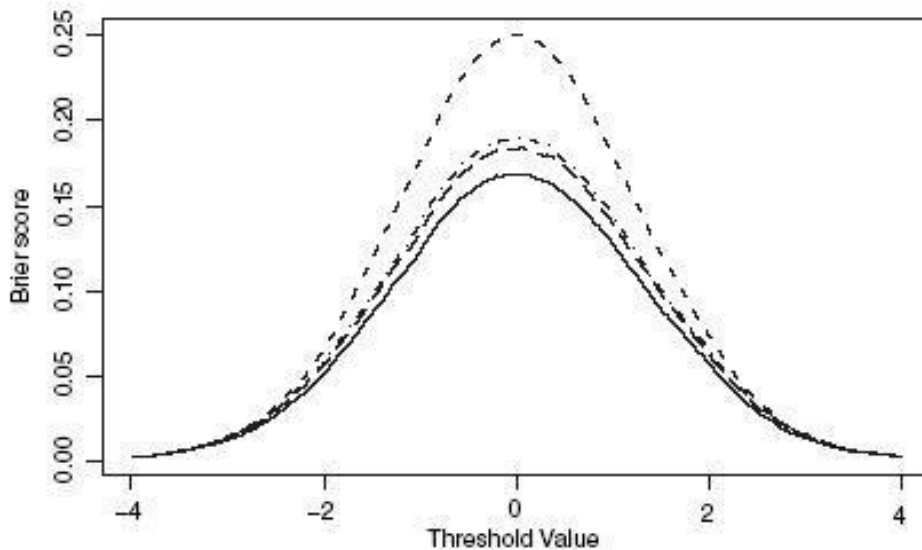
Most vizsgáljuk meg, hogy a logaritmikus szkór és a CRPS használatával milyen sorrendet tudunk felállítani az előrejelzések között. A 10.000 elemszámú minta alapján felírt átlag-szkórokat a következő táblázat foglalja össze:

<i>Előrejelző</i>	<i>LogS</i>	<i>CRPS</i>
Ideális	1.39	0.54
Klimatológiai	1.77	0.77
Nem-fókuszált	1.54	0.64
Hamill	1.53	0.61

2.3. táblázat. Átlagos szkór

Leolvashatjuk, hogy mindkét szkór az ideális előrejelzőt teszi meg a legjobbnak, ezután a Hamill és a nem-fókuszált előrejelző következik szorosan egymás után, majd utolsóként a klimatológiai előrejelző. A Hamill előrejelzőről beláttuk, hogy semelyik megismert kalibráció feltételének nem tesz eleget, mégis mindkét szkór előrébb sorolta a klimatológiai előrejelzőnél, pedig láttuk, hogy ő valószínűségi és marginálisan is kalibrált. Ez azért van így, mert bár se nem valószínűségi, se nem marginálisan kalibrált a Hamill-előrejelző, a (2.8) és a (2.9) formulákból leolvashatjuk, hogy csak egy kicsit sérti meg mindegyik feltételt, így igen közel van mindkét kalibráltsághoz.

A következő ábrán a $\{X < y\}$ -re felírt $BS(y)$ bináris Brier-szkórt⁸ láthatjuk az y küszöbérték függvényeként. Az előző fejezetben látottak szerint $CRPS = \int BS(y)dy$, tehát a görbék alatti területek épp a CRPS megfelelő értékét adják.



2.2. ábra. A Brier szkór és a CRPS kapcsolata

— ideális, - - - klimatológiai, -.-.- nem-fókuszált, . . . Hamill-előrejelző

⁸ $BS(y) = -(F_X(y) - \mathbf{1}\{X < y\})^2$

Diszkrét előrejelzések esete

Abban az esetben, ha a prediktív eloszlás diszkrét, akkor a PIT-értékek nem lesznek már egyenletesek az ideális előrejelző hipotézise esetében sem. Ezt orvoslandó sokan a következő *randomizált PIT*-et javasolják: legyen P a prediktív eloszlás és legyen x P eloszlású véletlen egész, valamint v legyen x -től független $(0, 1)$ -en egyenletes eloszlású szám. Ekkor

$$u = P_{x-1} + v(P_x - P_{x-1}), \quad x \geq 1 \quad (2.10)$$

$$u = vP_0, \quad x = 0$$

szintén egyenletes eloszlású $(0, 1)$ -n.⁹

Czado, Gneiting és Held [5]-ben a PIT-hisztogramm következő nem-randomizált, mégis egyenletes verzióját ismerteti: ehhez a randomizált PIT-értékeket (2.11)-ben helyettesítsük a feltételes eloszlásfüggvényekkel adott x megfigyelés mellett:

$$F(u) = \begin{cases} 0, & u \leq P_{x-1} \\ \frac{u - P_{x-1}}{P_x - P_{x-1}}, & P_{x-1} \leq u \leq P_x \\ 1, & u \geq P_x \end{cases} \quad (2.11)$$

ha $x \geq 1$, és

$$F(u) = \begin{cases} \frac{u}{P_0}, & u \leq P_0 \\ 1, & u \geq P_0 \end{cases}$$

ha $x = 0$.

A valószínűségi kalibráció ezután becsülhető ezek előrejelzések során kapott átlagát

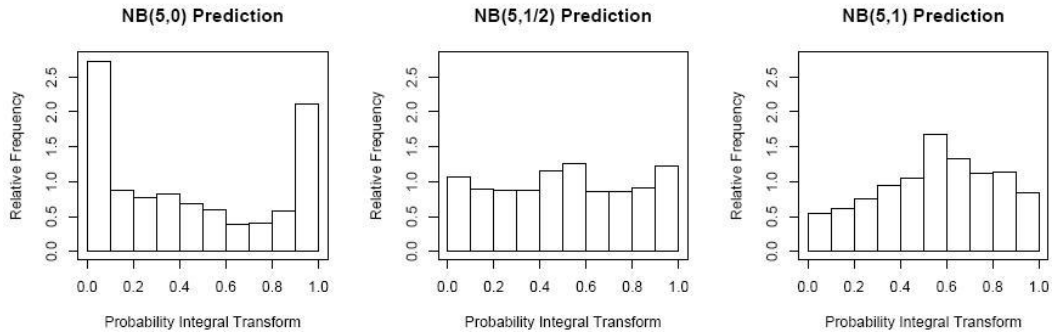
$$\bar{F}(u) = \frac{1}{n} \sum_{i=1}^n F^{(i)}(u), \quad 0 \leq u \leq 1 \quad (2.12)$$

összevetve az egyenletes eloszlás eloszlásfüggvényével (azaz az identitással). Itt $F^{(i)}$ a megfelelő $P^{(i)}$ prediktív eloszlás és az x_i megfigyelés alapján készített eloszlásfüggvény. Emellett felrajzolhatjuk a nem-randomizált PIT-hisztogrammot: osszuk fel $(0, 1)$ -et N egyenlő részre, majd számoljuk ki az $f_n = \bar{F}\left(\frac{n}{N}\right) - \bar{F}\left(\frac{n-1}{N}\right)$ értékeket és rajzoljuk fel a hisztogrammot a j . beosztásnál f_j súllyal, ezután ellenőrizhetjük az egyenletességet. Fontos megjegyezni, hogy a PIT-hisztogramm egyenletessége igen erős megkötés, és azzal ekvivalens, hogy minden prediktív intervallum nominális lefedettséget mutat. Ezért a néha igen hasznos módszer, miszerint táblázat foglalják néhány prediktív intervallum tapasztalati lefedettségét, a PIT-hisztogramm speciális eseteként interpretálható. Vegyünk egy prediktív intervallumot α alsó és β felső határral, ennek nominális lefedettsége $\beta - \alpha$. Randomizálással a tapasztalati lefedettség a randomizált PIT-értékek $[\alpha, \beta]$ intervallumba eső relatív gyakorisága, a nem-randomizált esetben pedig $\bar{F}(\beta) - \bar{F}(\alpha)$ -ként számolható. Természetesen ez a különbség a randomizált PIT alapján számított empirikus lefedettség várható értéke. A PIT-értékek hisztogramma nagy mintaelemszám esetén kihangsúlyozza az egyenletességtől való apró eltéréseket is, és ezek kimutatásához általában 10 vagy 20 beosztás elegendő.

⁹Itt $P_k = \sum_{i=0}^k p_i$.

Esettanulmány

Paraméterezzük a negatív binomiális eloszlást a következőképpen: $NB(\lambda, a)$ úgy, hogy a várható értéke $\lambda \geq 0$ legyen, míg a variancia $\lambda(1 + a\lambda) \geq 0$. Ekkor $a \geq 0$ és az $a = 0$ esetben a Poisson-eloszlást kapjuk speciális esetként. Egy 200 elemű mintát veszünk a $NB(5, \frac{1}{2})$ eloszlásból és a következő három előrejelzőt vizsgáljuk: $NB(5, 0) = P(5)$, $NB(5, \frac{1}{2})$ és $NB(5, 1)$. A (2.3) ábra a nem-randomizált PIT-hisztogramot mutatja 10 beosztással. Az első esetben homorú diagramot kapunk, ami a szórás alulbecslésére utal, azaz túl "szűk" a prediktív eloszlás; a második eset a valóságnak megfelelő; míg a harmadik esetben épp fordítva a domború diagram a szórás túlbecsléséből, tehát egy túl "széles" prediktív eloszlásból származik.



2.3. ábra. PIT-hisztogram

Általánosság megszorítása nélkül ez általánosítható arra az esetre is amikor nem feltétlen egészértékű, de rendezett adatsorral van dolgunk. Tehát egy x mennyiség előrejelzésével foglalkozunk, amely az $(x_k)_{k=-\infty}^{\infty}$ valós számokat veheti fel $(p_k)_{k=-\infty}^{\infty}$ valószínűséggel és $x_{k-1} < x_k < x_{k+1}$. $(P_k)_{k=-\infty}^{\infty}$ -t az előzőekhez hasonlóan képezzük. A PIT-transzformáció általánosítható erre az esetre is. Először nézzük a randomizált verziót: ha $x = x_k$ a megfigyelés:

$$u = P_{k-1} + v(P_k - P_{k-1}) \quad (2.13)$$

ahol v x -től független standard egyenletes eloszlású változó. A nem-randomizált esetben, $x = x_k$ megfigyelés mellett

$$F(u) = \begin{cases} 0, & u \leq P_{k-1} \\ \frac{u - P_{k-1}}{P_k - P_{k-1}}, & P_{k-1} \leq u \leq P_k \\ 1, & u \geq P_k \end{cases} \quad (2.14)$$

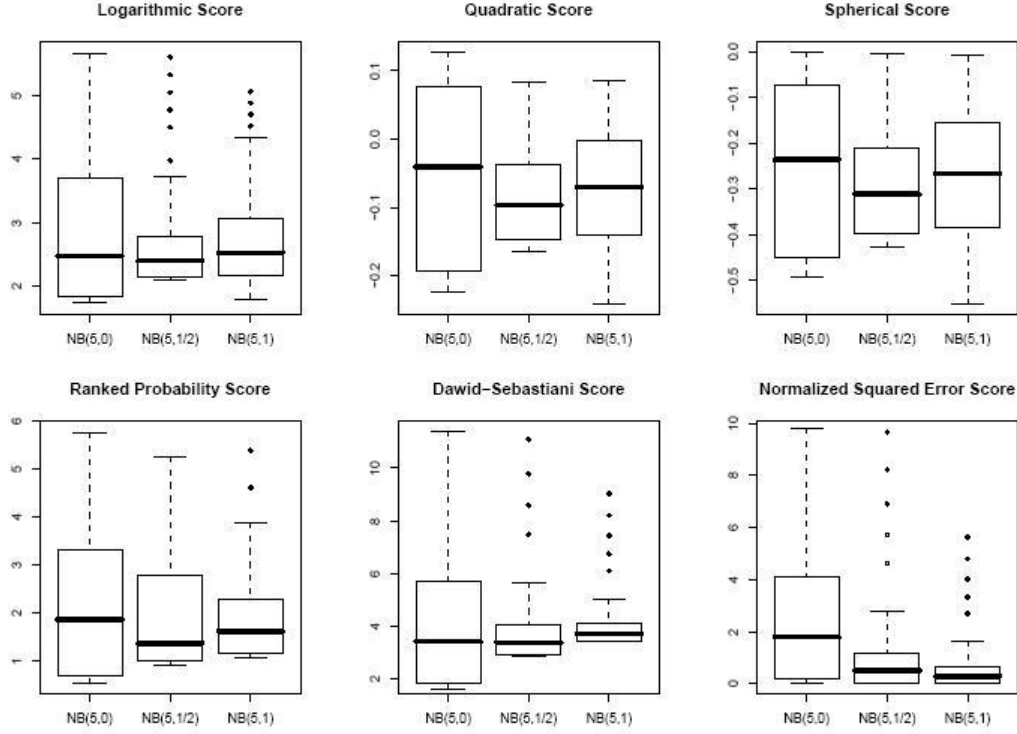
Melyek a nemnegatív egész értéket felvevő változók esetében a már ismert képleteket adják. Ismét aggregálunk, és a hisztogram egyenletességét vizsgálhatjuk.

Mint azt a következő

$$G_t \circ F_t^{-1}(p) = P(X_t < F_t^{-1}(p)) = P(F_t(X_t) < p) = P(p_t < p) \quad (2.15)$$

egyszerű átalakítás is mutatja, a valószínűségi kalibráltság ekvivalens a PIT-értékek egyenletességével.

A (2.4) ábrán a 200 mintaelemre kiszámolt szkórok tapasztalati eloszlását látjuk boxplot módszerrel ábrázolva. A középső vastag vonalak a mediánt, a dobozok határai az alsó és felső kvartilist¹⁰, míg az alsó és felső vonalak a legkisebb és legnagyobb szkórértékeket jelölik, ha ezek különbsége nem nagyobb a doboz magasságának 1,5-szeresénél. Ha ezen a határon kívülre is esnek megfigyelések, azokat a kívülálló pontok jelölik.



2.4. ábra. Boxplot

Nyilván a G_t eloszlása a valóságban csak hipotetikus, a gyakorlatban nem ismert, ezért jó volna, ha lenne egy olyan eszköz a kezünkben, mellyel a valódi háttéreloszlás ismerete nélkül is ellenőrizni tudnánk a valószínűségi kalibrációt. Persze G_t -t helyettesíthetjük az $\mathbf{1}\{x_t \leq x\}$ tapasztalati eloszlásfüggvénnyel, és vizsgálhatjuk az így kapott mennyiséget, következő tétel e segítségével karakterizálja a valószínűségi kalibrációt:

Tétel ([3]): Legyenek $(F_t)_{t=1,2,\dots}$ és $(G_t)_{t=1,2,\dots}$ folytonos és szigorúan monoton eloszlásfüggvények sorozatai. Legyen x_t eloszlásfüggvénye G_t és (x_t) *-keverő sorozat. Ekkor

$$\frac{1}{T} \sum_{t=1}^T \mathbf{1}\{p_t \leq p\} \rightarrow p \quad 1 \text{ valószínűséggel minden } p\text{-re} \quad (2.16)$$

akkor és csak akkor teljesül, ha az $(F_t)_{t=1,2,\dots}$ sorozat valószínűségi kalibrált a $(G_t)_{t=1,2,\dots}$ sorozathoz.¹¹

¹⁰Ezek a $\frac{1}{4}$ és a $\frac{3}{4}$ -szintű kvantilisok.

¹¹ (x_n) *-keverő sorozat, ha létezik N pozitív egész, és egy f valós értékű függvény, mely $n \geq N$ egészekre értelmezett a következő két tulajdonsággal:

(1) f monoton csökkenő és $\lim_{n \rightarrow \infty} f(n) = 0$

(2) ha $n \geq N$ és $A \in \sum_1^n$ illetve $B \in \sum_{m+n}$, akkor $|P(AB) - P(A)P(B)| \leq f(n)P(A)P(B)$

ahol \sum_n a legszűkebb σ -algebra, amire x_n mérhető és \sum_m^n a legszűkebb σ -algebra, amire x_m, \dots, x_n mérhetőek.

Kihangsúlyozottan: a (2.16) feltétel szükséges és elégséges a valószínűségi kalibrációhoz függetlenül attól, hogy mint (2.1) empirikus megfelelőjeként vezettük be.

Egy hasonló karakterizáció adható a marginális kalibrációra is, ahogy azt a következő részben látni is fogjuk; az exceedance kalibráció esetében viszont nem ismert hasonló módszer, mely (2.2) tapasztalati megfelelőjével karakterizálná az exceedance kalibrációt.

2.1.2. Marginális kalibráció becslése

Marginális kalibráció diagramm

Kézenfekvő ötlet, hogy ha minden egyes megfigyelés egy véletlen szám a megfelelő prediktív eloszlásból, akkor azt várnánk, hogy ha összesítjük ezeket a prediktív eloszlásokat, akkor az így kapott keverék eloszlás és a megfigyelt adatok hisztogramja statisztikailag összeegyeztethetőek. A marginális kalibráció diagramm egy x érték vagy egy $(x_a, x_b]$ intervallum **átlagos** valószínűségét hasonlítja össze az x megfigyelés empirikus gyakoriságával vagy az $(x_a, x_b]$ intervallum empirikus lefedettségével.

A marginális kalibráció (2.3) feltételében G_t sorozat ismételten a valóságban csak hipotetikus, a gyakorlatban nem ismert, ezért most is jó volna, ha lenne egy olyan eszköz a kezünkben, mellyel a valódi háttéreloszlás ismerete nélkül is ellenőrizni tudnánk a marginális kalibrációt. Most is helyettesíthetjük G_t -t az $\mathbf{1}\{x_t \leq x\}$ tapasztalati eloszlásfüggvénnyel, és vizsgálhatjuk $\widehat{G}_T(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}\{x_t \leq x\}$ és \overline{F}_T kapcsolatát.

Ezért is fontos a következő tétel, mely szerint enyhe regularitási feltételek mellett a marginális kalibráció szükséges és elegendő feltétele \widehat{G}_T és \overline{F}_T aszimptotikus egyenlőségének.

Tétel ([3]): *Legyenek $(F_t)_{t=1,2,\dots}$ és $(G_t)_{t=1,2,\dots}$ folytonos és szigorúan monoton eloszlásfüggvények sorozatai. Legyen x_t eloszlásfüggvénye G_t és (x_t) *-keverő sorozat. Továbbá tegyük fel, hogy*

$$\overline{F}(x) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{T} \sum_{t=1}^T F_t(x) \right\} \quad (2.17)$$

határérték létezik minden $x \in \mathbb{R}$ esetén és a határérték is szigorúan monoton. Ekkor

$$\widehat{G}_T(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}\{x_t \leq x\} \rightarrow \overline{F}(x) \quad 1 \text{ valószínűséggel minden } x\text{-re} \quad (2.18)$$

akkor és csak akkor teljesül, ha az $(F_t)_{t=1,2,\dots}$ sorozat marginálisan kalibrált a $(G_t)_{t=1,2,\dots}$ sorozathoz.

Így a marginális kalibráció ellenőrzése alapulhat a prediktív eloszlásfüggvények átlagának

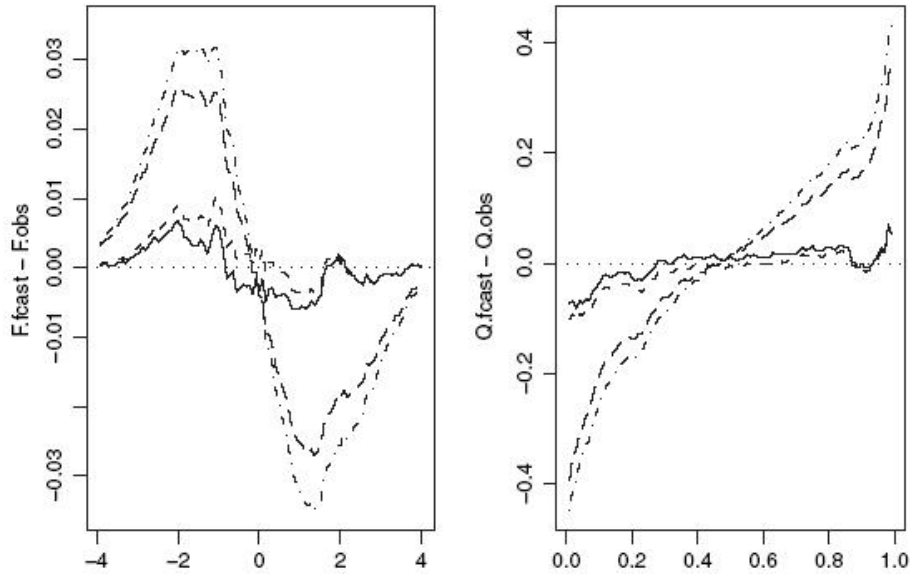
$$\overline{F}_T = \frac{1}{T} \sum_{t=1}^T F_t(x) \quad x \in \mathbb{R} \quad (2.19)$$

a megfigyelések tapasztalati eloszlásfüggvényével

$$\widehat{G}_T(x) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}\{x_t \leq x\} \quad x \in \mathbb{R} \quad (2.20)$$

való összevetésén.

A legegyszerűbb grafikus technika az $\bar{F}_T(x)$ és a $\hat{G}_T(x)$ függvények ábrázolása. Ennél sokkal informatívabb az $\bar{F}_T(x) - \hat{G}_T(x)$ különbség ábrázolása¹² (első ábra):



2.5. ábra. Marginális kalibráció diagramm az eloszlás- és a kvantilisfüggvények esetében, — az ideális, ---- a klimatológiai, a nemfókuszált, - - - pedig a Hamill előrejelzőt jelzi

A marginális kalibráció hipotézise mellett csak kis ingadozásokat várunk 0 körül, és ez teljesül is az ideális és a klimatológiai előrejelzőre. Mivel a másik két előrejelző nem marginálisan kalibrált, így nagy kilengéseket tapasztalunk az ábrán. Hasonló olvasható le a másik ábráról is, ahol a két kvantilisfüggvény $Q(\bar{F}_T, q) - Q(\hat{G}_T, q)$ különbsége látható. Megintcsak marginális kalibráció esetén 0 körüli kis fluktuációkat várunk, amely teljesül is az ideális és a klimatológiai előrejelző esetében. A másik kettő előrejelzőnél a kvantilisfüggvények különbsége egyre nő a negatív értékek felől a pozitívok fele, túl szétterjedt előrejelzett klímára utalva.

2.2. Élesség

Mint már említettem Gneiting, Raftery és mások a valószínűségi előrejelzések célját a következőképpen fogalmazzák meg: maximalizálni az eloszlás élességét rögzített kalibráció mellett. Ez lényegében az ideális előrejelzést jelenti, például a teljes erős kalibráltság megléte biztosítja az aszimptotikus ekvivalenciát. Tehát a bevezetett kalibrációfogalmaktól, azt várnánk, hogy egy megfelelően elégségesen kalibrált prediktív eloszlás legalább annyira szétterjedt mint az ideális előrejelzés. Az erős kalibráltság a teljesség nélkül nem tűnik elég erős megkötésnek, bár nem tudni ellenpéldáról. Azt viszont könnyen be lehet látni, hogy se a valószínűségi, se az exceedance, se a marginális kalibráció önmagában nem elég megkötés, azaz létezik előrejelző, mely valószínűségi, exceedance vagy marginálisan kalibrált, mégis élesebb az ideális előrejelzőnél.

¹²Az ábrán $F.\text{cast}=\bar{F}_T(x)$, $F.\text{obs}=\hat{G}_T(x)$, $Q.\text{cast}=Q(\bar{F}_T, q)$ és $Q.\text{obs}=Q(\hat{G}_T, q)$.

A következő példákban elég lesz a valószínűségi kalibráció véges $(F_t)_{1 \leq t \leq T}$ és $(G_t)_{1 \leq t \leq T}$ sorozatokra vonatkozó megfelelőjével foglalkoznunk:

$$\frac{1}{T} \sum_{t=1}^T G_t \circ F_t^{-1}(p) = p \quad \text{minden } p \in (0, 1) \quad (2.21)$$

Hasonlóan definiálható a többi kalibráció is véges sorozatokra, például az exceedance kalibrációra:

$$\frac{1}{T} \sum_{t=1}^T G_t^{-1} \circ F_t(x) = x \quad \text{minden } x \in \mathbb{R} \quad (2.22)$$

Legyen $\sigma > 0$, $a > 1$, $0 < \lambda < \frac{1}{a}$ és $T = 2$, illetve legyenek G_1 és G_2 folytonos, szigorúan monoton eloszlásfüggvények, 0-ra szimmetrikus sűrűségfüggvényekkel és véges szórásnégyzetekkel: $D^2(G_1) = \sigma^2$ és $D^2(G_2) = \lambda\sigma^2$.

Definiáljuk ekkor F_1 -t és F_2 -t a következőképpen:

$$F_1(x) = \frac{1}{2} \left\{ G_1(x) + G_2\left(\frac{x}{a}\right) \right\}$$

$$F_2(x) = F_1(ax)$$

Ekkor ezek varianciájára a következő igaz:

$$D^2(F_1) + D^2(F_2) = \frac{1}{2} \left(1 + \frac{1}{a^2} \right) (1 + a^2\lambda^2)\sigma^2 < (1 + \lambda^2)\sigma^2 = D^2(G_1) + D^2(G_2)$$

jöllehet a véges valószínűségi kalibráció (2.21) feltétele teljesül.

Tehát egy valószínűségi kalibrált előrejelző lehet élesebb, mint az ideális előrejelző.

Hasonló példa adható a exceedance kalibrációra is:

legyen $\sigma > 0$ és $0 < a < 1$, és $0 < \lambda < a \left(\frac{3+a}{1+3a} \right)^{1/2}$, G_1 és G_2 ugyanolyan mint előbb és definiáljuk F_1 -et és F_2 -öt a következőképpen:

$$F_1(x) = G_1\left(\frac{2x}{1+a}\right)$$

$$F_2(x) = G_2\left(\frac{2ax}{1+a}\right)$$

Ekkor ezek varianciájára:

$$D^2(F_1) + D^2(F_2) = \frac{1}{4}(1+a)^2 \left(1 + \frac{\lambda^2}{a^2} \right) \sigma^2 < (1 + \lambda^2)\sigma^2 = D^2(G_1) + D^2(G_2)$$

noha a véges exceedance kalibráció feltétele megintcsak teljesül.

Tehát egy exceedance kalibrált előrejelző is lehet élesebb az ideális előrejelzőnél. Nyilván egy marginálisan kalibrált előrejelző is lehet élesebb mint az ideális előrejelző.

Klimatológiai előrejelzések esetében a véges valószínűségi és a véges marginális kalibráltság ekvivalens és egy gyenge élességi állítás is igaz, nevezetesen alsó határ biztosítható a prediktív eloszlás varianciájára:

Tétel [3]: Véges második momentummal rendelkező G_1, \dots, G_T és $F_1 = \dots = F_T = F$ esetén, ha a véges valószínűségi kalibráció (2.21) feltétele teljesül, akkor

$$\frac{1}{T} \sum_{t=1}^T D^2(F_t) = D^2(F) \geq \frac{1}{T} \sum_{t=1}^T D^2(G_t) \quad (2.23)$$

és egyenlőség csak az $E(G_1) = \dots = E(G_T)$ esetben lehetséges.

Nyitott kérdés, hogy vajon egy nem-klimatológiai előrejelző lehet-e valószínűségi és marginálisan kalibrált, mégis élesebb mint az ideális előrejelző.

2.2.1. Az élesség becslése

Az előző példákban a prediktív eloszlások szórásnégyzetével mértük az élesség mértékét, most egy másik eszközt mutatok be, mellyel ez szintén megtehető. Ez a prediktív intervallumok szélességének vizsgálata, melyre két lehetőség is adódik. Az egyik a prediktív intervallumok átlagos hosszának vizsgálata, a másik pedig egy grafikus eszköz, mely a prediktív intervallumok szélességének tapasztalati eloszlását ábrázolja boxplottal.

A következő táblázat az 50%-os és a 90%-os prediktív intervallumok átlagos hosszát foglalja össze a 4 eddig is vizsgált előrejelző esetén.

Előrejelző	Átlagos hossz	
	50%	90%
Ideális	1.33	3.35
Klimatológiai	1.91	4.65
Nem-fókuszált	1.55	3.78
Hamill	1.46	3.59

2.4. táblázat. Az 50% és a 90%-os prediktív intervallumok átlagos hossza

Láthatjuk, hogy mindegyik esetben az ideális előrejelző esetében a legrövidebb a prediktív intervallumok átlagos szélessége, ezután a Hamill-előrejelző következik, majd a nem-fókuszált és végül a klimatológiai előrejelző.

3. fejezet

Bayesi megközelítés

A bayesi hozzáállás kezdeti feltevése valamilyen a priori ismeret a paraméter(ek)ről, valamilyen $p(\theta)$ a priori eloszlás formájában. A Bayes-tétel és a rendelkezésre álló adatok segítségével ilyenkor kiszámítható a paraméter(ek) a posteriori eloszlását:

$$p(\theta|x) = \frac{1}{p(x)} \cdot p(x|\theta) \cdot p(\theta) = \text{normalizáló konstans} \cdot \text{likelihood} \cdot \text{prior} \quad (3.1)$$

ugyanis $p(x|\theta) = l(\theta|x)$ a likelihood-függvény és $\frac{1}{p(x)}$ konstansként viselkedik a θ változó mellett, ugyanis a paraméter(ek) eloszlására vagyunk kíváncsiak. Ezért a posteriori eloszlást gyakran a

$$p(\theta|x) \propto l(\theta|x)p(\theta) \quad (3.2)$$

alakban írják, ahol \propto szimbólum jelentése: "arányos vele", vagy "multiplikatív konstans erejéig egyenlőek". A $p(x)$ konstans normalizálja $p(x|\theta)p(\theta)$ -t eloszlássá, így megkapható a következőből

$$p(x) = \int_{\Theta} p(x|\theta)p(\theta)d\theta \quad (3.3)$$

Az a posteriori eloszlás a prioritól való függése mutatja mennyi információt hordoz az adathalmaz az ismeretlen paraméter(ek)ről. Ezt letehetjük, hogy ha különböző priorokat választunk és megnézzük, hogy az a posteriori eloszlások mennyire térnek el egymástól. Néha csak a paraméterek egy része érdekel minket, a többi ún. zavaró paraméter nem. A bayesi analízis nagy érdeme, hogy ki tudjuk szűrni vele az ilyen zavaró paraméterek hatását a többi paraméter becslésénél, egyszerűen integrálással. Az így kapott marginális a posteriori eloszlás a főbb paraméterekről hordoz csak információt. Például képzeljük el, hogy egy normális eloszlás mindkét paramétere ismeretlen számunkra, de csak a variancia érdekel igazán minket. A várható érték becslése plusz bizonytalanságot ad a variancia becsléséhez. Ekkor a variancia marginális a posteriori eloszlása a következőképpen számolható:

$$p(\sigma^2|x) = \int p(\mu, \sigma^2|x)d\mu \quad (3.4)$$

Általánosan pedig írjuk a paraméterek vektorát a $\theta = (\theta_1, \theta_2)$ alakban, ahol θ_2 a zavaró paraméterek részvektora, ekkor a marginális a posteriori eloszlásunk:

$$p(\theta_1|x) = \int_{\Theta_2} p(\theta_1, \theta_2|x)d\theta_2 \quad (3.5)$$

Hogy kaphatunk konkrét pontbecsléseket az ismeretlen θ paraméterre? Vehetjük a posteriori eloszlás valamilyen funkcionálját, például a maximum likelihood terminológiával vehetjük a

posteriori eloszlás móduszát, azaz a legvalószínűbb értéket

$$\hat{\theta} = \max_{\theta} p(\theta|x) \quad (3.6)$$

Vagy vehetjük θ a posteriori várható értékét:

$$\hat{\theta} = E(\theta|x) = \int \theta p(\theta|x) d\theta \quad (3.7)$$

Esetleg a posteriori eloszlás mediánja is szóba jöhet, ami a következőt teljesíti:

$$\int_{\hat{\theta}}^{+\infty} p(\theta|x) d\theta = \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta = \frac{1}{2} \quad (3.8)$$

Habár ezen becslések valamelyikét, vagy akár mindegyiket szimultán használva a Bayesi analízis a fő erejét veszti el, mely abban rejlik, hogy **a becslése az egész a posteriori eloszlás**. Gyakran a posteriori eloszlást Gibbs mintavételezéssel kapják meg.

Legnagyobb sűrűségű régiók

Adott a posteriori eloszlás mellett persze kézenfekvő megfelelő lefedettségű konfidencia-intervallumok konstruálása. Például egy $100 \times (1 - \alpha)\%$ megbízhatósági szintű konfidencia-intervallum minden olyan (L_{α}, R_{α}) pár, mely kielégíti a következő egyenletet:

$$\int_{L_{\alpha}}^{R_{\alpha}} p(\theta|x) d\theta = 1 - \alpha \quad (3.9)$$

Szűkítsük le a szóbjövő intervallumokat azokra, melyek hossza a legrövidebb (több paraméter esetén a területe), ezek a Bayesi konfidenciaintervallumok.¹

3.1. Konjugált priorok

A bayesi analízis legelső feladata a megfelelő prior kiválasztása, megtalálása. Az egyik elterjedt prior, mikor a sűrűségfüggvény egy konstans, azaz a priori eloszlás egyenletes:

$$p(\theta) = c = \frac{1}{b - a} \quad a \leq \theta \leq b \quad (3.10)$$

Ez annak a gondolatnak felel meg, hogy nincs semmi okunk megkülönböztetni a paraméterértékeket, mindegyiket egyforma valószínűségűnek érezzük. Ekkor az a posteriori eloszlás csak egy konstansszorosa a likelihoodnak:

$$p(\theta|x) \propto l(\theta|x) \quad (3.11)$$

Sok esetben egy-egy klasszikus eredményt úgy kapunk meg a bayesi analízis segítségével, hogyha ilyen priort feltételezünk. Persze ez csak akkor működik, hogyha a paraméterértékünk véges határok között mozog. Na de mi a helyzet más esetekben, hogyan válasszuk ekkor meg az a priori eloszlást?

Az egész kulcsa az, hogy ha az adathalmazban elég információ van, akkor még egy rossz prior se fogja befolyásolni a posteriori viselkedést. A bayesi analízis aszimptotikus tulajdonsága, hogy patológikus prioroktól eltekintve, elegendő mennyiségű adatot véve, az egy idő után már megfelelő mennyiségű információt hordoz magával, s így az a posteriori eloszlás

¹A *credible intervals* és a *highest density regions (HDRs)* elnevezések is használatosak az angol irodalomban.

nem függ a priortól lényeges módon, így tetszőleges, a feladatnak megfelelő (nem elfajult) eloszlást vehetünk priorinak. Továbbra is fennáll a kérdés: mennyi az a megfelelő adat? Ezt például leellenőrizhetjük úgy, hogy különböző priorokat veszünk, és megnézzük mennyire befolyásolja ez az adódó posteriori eloszlást. Ha a posteriori eloszlások nagy mértékben függenek a különböző priori választásoktól, akkor az adathalmaz nem tartalmaz elegendő információt. Ha viszont a posteriori elég stabilnak tűnik különböző priori eloszlások esetén, akkor az adathalmaz által hordozott információ elegendő.

Általában az a priori eloszlás paraméterei (várható értéke illetve varianciája) sokkal inkább kritikusabb szerepet játszanak, mint az adott prior alakja az a posteriori viselkedésnél. Erre majd később példát is láthatunk Lindley nevezetes paradoxonánál. Ezért az a priori eloszláscsaládot általában úgy választják, hogy megkönnyítse az a posteriori eloszlás számítását. Például Poisson likelihood mellett Gamma-eloszlást választva priorinak, könnyen kiszámolható, hogy posterioriként szintén Gamma-eloszlást kapunk (a megfigyelésekkel transzformált paraméterekkel). Ilyen *konjugált priorok* más esetben is léteznek, mikor adott likelihood mellett a posterior eloszláscsaládja megegyezik a prior eloszláscsaládjával. Megmutatható, hogy ha a likelihood az exponenciális család tagja, akkor mindig található (szintén az exponenciális családba tartozó) konjugált prior hozzá. Ehhez legyen a likelihood a következő alakú:

$$l(x_i|\theta) = g(\theta)h(x_i) \exp\left(\sum_{j=1}^m \phi_j(\theta)t_j(x_i)\right) \quad (3.12)$$

Ekkor a következő priorral

$$p(\theta) \propto [g(\theta)]^b \exp\left(\sum_{j=1}^m \phi_j(\theta)a_j\right) \quad (3.13)$$

az a posteriori eloszlásunk

$$p(\theta|x) \propto \left[\prod_{i=1}^n l(x_i|\theta)\right] p(\theta) \propto [g(\theta)]^{b+n} \exp\left(\sum_{j=1}^m \phi_j(\theta)d_j(x)\right) \quad (3.14)$$

ahol

$$d_j(x) = a_j + \sum_{i=1}^n t_j(x_i) \quad (3.15)$$

A b és az a_j paraméterek pedig szabadon választhatók például a priorok momentumainak előírásához. A likelihood és a prior sűrűségfüggvénye nagyon hasonlít egymáshoz, de más a változó, az elsőben $\phi_j(\theta)$ mint paraméterezés tűnik fel, addig a priornál mint elégséges statisztika.

Tehát a likelihood-függvényben a $\phi_j(\theta)$ paraméterezésnek megfelelő exponenciális családbeli prior felírása az első feladatunk, majd a b és az a_j ún. *hiperparaméterek* meghatározása. Egyszerűen adódik, hogy ugyanabba az exponenciális családba tartozik a posteriori eloszlásunk, mint a priori. Az adatok beérkezésének hatására bekövetkező változást a posteriori paraméterekre is könnyen leolvashatjuk a (3.13) képletből: b paraméterhez az n mintaelem-szám adódik, míg az a_j -k változását a (3.14)-ből olvashatjuk le és láthatjuk, hogy ez csak a $(\sum_{i=1}^n t_j(x_i))_{j=1}^m$ elégséges statisztikától függ. Ez azért van így, mert a likelihood-függvény is csak az elégséges statisztikán keresztül függ a megfigyelésektől: két adathalmaz esetén, melyek megfigyelései különbözhetnek, de az elégséges statisztikájuk megegyezik, akkor a likelihood-függvénye is megegyezik a két adathalmaznak.

Likelihood	Konjugált prior
Geometriai	Béta
Negatív binomiális	Béta
Binomiális	Béta
Poisson	Gamma
Exponenciális	Gamma
Gamma	Gamma
Normális	
μ ismeretlen, σ^2 ismert	Normális
μ ismert, σ^2 ismeretlen	Inverz- χ^2

3.1. táblázat. **Konjugált priorok**

Tehát minden exponenciális családbeli likelihoodhoz létezik egy szintén exponenciális eloszláscsaládú priori eloszlás, úgy hogy a posteriori eloszlás ugyanannak az eloszláscsaládnak a tagja, mint a priori. Ezt a priort nevezzük az adott likelihoodhoz tartozó *konjugált prior*nak.

A (3.1) táblázatban néhány ismert likelihoodhoz tartozó konjugált priorok találhatóak, leolvasható, hogy több likelihoodhoz is tartozhat ugyanaz a prior, így a megfeleltetés a kettő között nem egy-egyértelmű. Nézzük meg az ismeretlen várható értékű, de ismert varianciájú normális eloszlás esetét. A likelihood-függvény ekkor

$$l(\mu|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad (3.16)$$

Legyen a $p(\mu)$ a priori eloszlás szintén normális μ_0 és σ_0^2 hiperparaméterekkel: $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, azaz

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right) \quad (3.17)$$

Ekkor az a posteriori eloszlás a (3.2) képlet alapján a priorból és a likelihoodból a következőképpen rakható össze:

$$p(\mu|x) \propto l(\mu|x)p(\mu) \propto \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad (3.18)$$

Ez átalakítás után (összegyűjtjük a μ -t tartalmazó tagokat, a többi meg konstansként viselkedik) a következő alakú lesz:

$$p(\mu|x) \propto \exp\left(-\frac{\mu^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0^2} + \frac{\mu n\bar{x}}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \quad (3.19)$$

Végül μ szerint teljes négyzetté alakítva:

$$p(\mu|x) \propto \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2} + f(x, \mu_0, \sigma_0^2, \sigma^2)\right) \quad (3.20)$$

ahol $\sigma_*^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$ és $\mu_* = \sigma_*^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right)$.

Tehát az a posteriori eloszlás szintén normális μ_* és σ_*^2 paraméterekkel.

Kiszámolható ismert várható érték mellett, a varianciára is a konjugált prior, ami jelen esetben inverz- χ^2 lesz (egy χ^2 eloszlású változó reciprokának az eloszlása).

3.2. Hipotézisvizsgálat és Bayes faktor

A klasszikus hipotézisvizsgálati feladatban két alternatíva áll előttünk, az egyik a H_0 nullhipotézis, miszerint az ismeretlen θ paraméter a $\Theta_0 \subseteq \Theta$ résztartományba esik a paraméterterén. A H_1 ellenhipotézis szerint pedig $\theta \in \Theta_1 \subseteq \Theta$, ahol Θ_0 és Θ_1 diszjunkt felbontását adják Θ -nak. Elsőre azt gondolhatnánk, hogy a hipotézisvizsgálat a bayesi környezetben egyszerűen kezelhető, hiszen nézhetjük a hipotézisek a posteriori valószínűségét:

$$P(\theta \in \Theta_0|x) = \int_{\Theta_0} p(\theta|x)d\theta \quad (3.21)$$

Viszont a bayesi analízis velejárója az a priori ismeret, mely befolyásolja a hipotézisvizsgálatot is. Képzeljük el, hogy az a priori eloszlás szerint $P(\theta \in \Theta_0) = 0,1$, míg az a posteriori szerint $P(\theta \in \Theta_0|x) = 0,05$. Ez utóbbi a 95% szignifikanciaszintnek felel meg a klasszikus környezetben, de a beérkező adat csak megduplázta bizalmunkat az alternatív hipotézisben. Ha például $P(\theta \in \Theta_0) = 0,5$ lett volna a priori információnk, akkor ez az 95%-os a posteriori szignifikancia igencsak megerősítené az ellenhipotézisbe vetett hitünket. Tehát a prior igencsak befolyásolja a bayesi hipotézisvizsgálatot. Ezért vezessük be az a priori és az a posteriori valószínűségeit a hipotéziseinknek:

$$\begin{aligned} p_0 &= P(\theta \in \Theta_0|x), & p_1 &= P(\theta \in \Theta_1|x) \\ \pi_0 &= P(\theta \in \Theta_0), & \pi_1 &= P(\theta \in \Theta_1) \end{aligned} \quad (3.22)$$

Definíció: H_0 a priori esélye H_1 ellenében π_0/π_1 , míg az a posteriori esélye p_0/p_1 .

A $B_0(x)$ Bayes faktort² a következőképp definiáljuk:

$$B_0(x) = \frac{p_0/p_1}{\pi_0/\pi_1} = \frac{p_0(1-\pi_0)}{\pi_0(1-p_0)} \quad (3.23)$$

Azaz az a posteriori és az a priori esélyek hányadosa, így a Bayes faktor azt méri, hogy a bejövő adatok vajon növelik vagy csökkentik az egyik modell esélyeit a másikkal szemben. Ha $B_0(x) > 1$, akkor x fényében a H_0 nullhipotézis még inkább elfogadható, mint a H_1 ellenhipotézis. A Bayes faktor egy laza interpretációja: H_0 -nak az adathalmaz által hordozott esélyei H_1 ellenében. Fontos, bár a jelölésben nem látszik, hogy a Bayes faktor az a priori eloszlástól is függ, néha igen kényelmetlenül.

A példában a B_1 Bayes faktor az első esetben 2.11, míg a második esetben 19. Vizsgáljuk meg mit kapunk a $\Theta_0 = \theta_0$, $\Theta_1 = \theta_1$ egyszerű hipotézisek esetében, ekkor az a posteriori valószínűség:

$$p_i \propto p(\theta_i)p(x|\theta_i) = \pi_i p(x|\theta_i) \quad (3.24)$$

Így az a posteriori esély

$$\frac{p_0}{p_1} = \frac{\pi_0 p(x|\theta_0)}{\pi_1 p(x|\theta_1)} \quad (3.25)$$

és a Bayes faktor egyszerűen a likelihood-hányados

$$B_0(x) = \frac{p(x|\theta_0)}{p(x|\theta_1)} \quad (3.26)$$

²Ez a H_0 -ra vonatkozó Bayes faktor (H_1 ellenében), a H_1 -re vonatkozó $B_1(x)$ ennek reciprokaként kapható.

Összetett hipotézisek esetén vegyük θ H_i -re vett feltételes a priori eloszlását:

$$p_i(\theta) = p(\theta)/\pi_i \quad (3.27)$$

Ekkor az a posteriori valószínűség a következőképpen számítható ki:

$$p_i = P(\theta \in \Theta_i | x) = \int_{\theta \in \Theta_i} p(\theta | x) d\theta \propto \int_{\theta \in \Theta_i} p(\theta) p(x | \theta) d\theta = \pi_i \int_{\theta \in \Theta_i} p_i(\theta) p(x | \theta) d\theta \quad (3.28)$$

Ekkor a Bayes faktor (az a priori ismereteinkkel) súlyozott likelihoodhányados:

$$B_0(x) = \frac{\int_{\theta \in \Theta_0} p_0(\theta) p(x | \theta) d\theta}{\int_{\theta \in \Theta_1} p_1(\theta) p(x | \theta) d\theta} \quad (3.29)$$

Természetesen felírható ez egyszerű nullhipotézis esete az összetett ellenhipotézissel szemben, ezt egy példán ismertetem, mely egyben rávilágít a módszer egy hibájára:

3.2.1. Lindley paradoxona [9]

A következő paradoxon mutatja, hogy a Bayes faktor néha nagyon is szélsőséges módon függ az a priori eloszlástól, ugyanis a hiperparaméterek megfelelő(en radikális) választásával esetleg elérhető, hogy a beérkező adatot figyelmen kívül hagyva mindig elfogadjuk a nullhipotézist.

Legyen $x = (x_1, \dots, x_n)$ $\mathcal{N}(\mu, \sigma^2)$ eloszlású minta, ahol σ^2 ismert, μ viszont ismeretlen paraméter, és a $H_0 : \mu = \mu_0$ egyszerű nullhipotézist teszteljük az összetett $H_1 : \mu \neq \mu_0$ ellenhipotézissel szemben. A $p(\mu)$ a priori eloszlást válasszuk normálisnak μ_1 és σ_1^2 hiperparaméterekkel, ugyanis a (3.1) táblázatból kiolvasható, hogy ez lesz a likelihoodnak megfelelő konjugált prior.

Ekkor a $B_0(x)$ Bayes faktor a következő alakot ölti:

$$B_0(x) = \frac{p(x | \mu_0)}{\int p(x | \mu) p(\mu) d\mu} = \frac{\prod_{i=1}^n \frac{1}{\sigma} \varphi\left(\frac{x_i - \mu_0}{\sigma}\right)}{\int \prod_{i=1}^n \frac{1}{\sigma} \varphi\left(\frac{x_i - \mu}{\sigma}\right) \frac{1}{\sigma_1} \varphi\left(\frac{\mu - \mu_1}{\sigma_1}\right) d\mu} \quad (3.30)$$

Ez megfelelő átalakítások után:

$$B_0(x) = \left(1 + n \frac{\sigma_1^2}{\sigma^2}\right)^{1/2} \cdot \frac{\exp\left\{\frac{1}{2}\left(\sigma_1^2 + \frac{\sigma^2}{n}\right)^{-1}(\bar{x} - \mu_1)^2\right\}}{\exp\left\{\frac{1}{2}\frac{n}{\sigma^2}(\bar{x} - \mu_0)^2\right\}} \quad (3.31)$$

Ha most lerögzítjük \bar{x} értékét és $\sigma_1 \rightarrow \infty$ esetén vizsgáljuk a kifejezést, akkor az szintén végtelenhez tart. Tehát a H_0 hipotézis elfogadása mellett fogunk dönteni ha μ a priori szórásnégyzete elég nagy, akármit is sugall az adathalmaz. Minden rögzített \bar{x} esetén σ_1^2 értékét meg tudjuk úgy választani, hogy a nullhipotézist elfogadjuk. Speciális ez igaz olyan \bar{x} -re is, melyre $\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$ elég nagy ahhoz, hogy elutasítsuk a nullhipotézist egy előre tet-szőlegesen megadott szignifikanciaszinten a megfelelő u -próbával.

Egy kompromisszumot a Bayesi és a klasszikus hipotézis vizsgálat között Lindley ([9], 1965) adott egyszerű null- és kétoldali ellenhipotézis esetén:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

Először írjunk elő θ -ra egy szétterjedt priort, ekkor egy α szintű teszt a következőképpen néz ki: vegyünk egy $100(1 - \alpha)\%$ -os Bayesi konfidenciaintervallumot és utasítsuk el a nullhipotézist, ha ez nem tartalmazza θ -t.

3.3. Bayes faktor az általános esetben

Eddig csak hipotézisvizsgálathoz használtuk a Bayes faktort, tehát adott likelihood esetén az igazi paraméterértéket kerestük. Definiálható a Bayes faktor általánosabban is, amikor nem tudjuk, hogy az eloszlásunk mely paraméteres családból származik, nem ismerjük a likelihoodot sem. Az előző részben a Bayes faktor az a posteriori és az a priori esélyek hányadosa volt, próbáljuk meg ezt átmenteni az általánosabb esetre is. Most egy modell megadásához meg kell adnunk a paraméter(ek) a priori eloszlását és a likelihoodfüggvényt is. Legyen $\{M_j : j = 1, 2\}$ két paraméteres modell (likelihood+prior).

Definiáljuk M_1 a posteriori esélyét M_2 ellenében a következő módon:

$$\frac{p(M_1|x)}{p(M_2|x)} = \frac{p(M_1)}{p(M_2)} \frac{p(x|M_1)}{p(x|M_2)} \quad (3.32)$$

$$\left(\begin{array}{c} \text{posteriori} \\ \text{esély} \end{array} \right) = \left(\begin{array}{c} \text{priori} \\ \text{esély} \end{array} \right) \left(\begin{array}{c} \text{Bayes} \\ \text{faktor} \end{array} \right)$$

Ez alapján a Bayes faktort a következőképpen definiálhatjuk:

$$B_1(x) = \frac{p(x|M_1)}{p(x|M_2)} = \frac{\int p_1(x|\theta_1)p_1(\theta_1)d\theta_1}{\int p_2(x|\theta_2)p_2(\theta_2)d\theta_2} \quad (3.33)$$

Itt tehát a likelihoodfüggvény és az a priori eloszlás is függ az M_j modelltől.

A következő példa [4]-ben található: legyen (x_1, \dots, x_n) egész-értékű adatsor, melyre a következő két összetett modell valamelyikét akarjuk illeszteni:

M_1 : (0-ba betölt) Geometriai(θ_1) likelihood Béta(α_1, β_1) priorral

M_2 : Poisson(θ_2) likelihood Gamma(α_2, β_2) priorral

Ekkor M_1 -re felírva a Bayes-faktort M_2 ellenében:

$$B_1(x) = \frac{\Gamma(\alpha_1 + \beta_1)\Gamma(\alpha_1 + n)\Gamma(n\bar{x} + \beta_1)\Gamma(\alpha_2)(n + \beta_2)^{n\bar{x} + \alpha_2}(\prod_{i=1}^n x_i!)}{\Gamma(\alpha_1)\Gamma(\beta_1)\Gamma(n + n\bar{x} + \beta_1 + \alpha_1)\Gamma(n\bar{x} + \alpha_2)\beta_2^{\alpha_2}} \quad (3.34)$$

Ha most szétkent priorokat választunk, azaz $(\alpha_1, \beta_1) = (1, 1)$ és $(\alpha_2, \beta_2) = (\epsilon, \epsilon)$ (ez esetben a Béta eloszlás az egyenletesre redukálódik), akkor a Bayes-faktor a következőre redukálódik:

$$B_1(x) = \frac{\Gamma(1 + n)\Gamma(n\bar{x} + 1)\Gamma(\epsilon)(n + \epsilon)^{n\bar{x} + \epsilon}(\prod_{i=1}^n x_i!)}{\Gamma(n + n\bar{x} + 2)\Gamma(n\bar{x} + \epsilon)\epsilon^\epsilon} \quad (3.35)$$

Ha most ϵ -nal nullához tartunk, akkor ez a kifejezés végtelenhez tart. Tehát a Geometriai modell az a priori hiperparaméterek megfelelő választása mellett minden esetben kizozható jobbnak a Poissonnal szemben, mint ahogy ezt már a Lindley-paradoxonnál is tapasztaltuk. Tehát a Bayes faktort nagymértékben befolyásolja az a priori hiperparamétereinek megválasztása.

Most vizsgáljuk meg ezt általánosan is, legyen két paraméteres modellünk ($j = 1, 2$)

$$M_j : (x_i|\theta_j, M_j) \sim p_j(x_i|\theta_j) = L_j(\theta_j|x_i) \text{ likelihooddal és } p_j(\theta_j) \text{ priorral} \quad (3.36)$$

A közönséges Bayes faktor a következő mennyiségek összehasonlításán alapszik:

$$p(x|M_j) = \int \left[\prod_{i=1}^n p_j(x_i|\theta_j) \right] p_j(\theta_j) d\theta_j = E_{(\theta_j|M_j)} L_j(\theta_j|x) \quad (3.37)$$

Azaz a likelihoodfüggvények a priori várható értékét veti össze, ezért is viselkedik olyan rosszul nem megfelelő prior esetén.

A $B_1(x)$ alapján akkor favorizáljuk M_1 -et, ha

$$\log p(x|M_1) > \log p(x|M_2) \quad (3.38)$$

Ebben az alakban első ránézésre úgy tűnhet, hogy ez ugyanaz mint a logaritmikus szkór alapján hozott döntésünk

$$n \overline{\text{LogS}}(M_1, X) > n \overline{\text{LogS}}(M_2, X) \quad (3.39)$$

de jobban megvizsgálva a dolgot mégsem, ugyanis:

$$n \overline{\text{LogS}}(M_j, X) = \log \prod_{i=1}^n \left[\int p_j(x_i|\theta_j) p_j(\theta_j) d\theta_j \right] \quad (3.40)$$

Az integrál és a produktum operátorok pedig nem kommutálnak.

Aitkin ([4], 1991) az a priori helyett a likelihoodfüggvények a posteriori várható értékeket hasonlítja össze. Az így kapott *posteriori Bayes faktor* (PBF) azt a modellt favorizálja, amelyekre a

$$\log \bar{L}_j^A = \log \int \left[\prod_{i=1}^n p_j(x_i|\theta_j) \right] p_j(\theta_j|x) d\theta_j \quad (3.41)$$

kifejezés nagyobb értéket vesz fel. Ez megoldja a priorra való érzékenységet, de egy újat is magával hoz, méghozzá inkoherens. Mindemellett többen a loglikelihoodok a posteriori várható értéken alapuló összehasonlítást is javasolják.

$$\begin{aligned} E_{(\theta_j|x, M_j)} \log L_j(\theta_j|x) &= E_{(\theta_j|x, M_j)} \log \prod_{i=1}^n L_j(\theta_j|x_i) \\ &= \sum_{i=1}^n E_{(\theta_j|x, M_j)} \log L_j(\theta_j|x_i) < \sum_{i=1}^n \log E_{(\theta_j|x, M_j)} L_j(\theta_j|x_i) = \\ &= \sum_{i=1}^n \log \int p_j(x_i|\theta_j) p_j(\theta_j|x) d\theta_j \end{aligned} \quad (3.42)$$

Ez utóbbi kifejezés felfogható mint az M_j -nek megfelelő likelihooddal és a paraméter a posteriori eloszlásával adott modellre felírt logaritmikus szkór. Ez egy afféle *a posteriori logaritmikus szkór*, mikor ilyen összetett (likelihood+prior) modellel van dolgunk. Ezt is használhatjuk a két modell közötti választásra, és látható, hogy nem ekvivalens az előzővel.

4. fejezet

A logaritmikus szkór határeloszlása néhány esetben

Klasszikus statisztikai feladatok esetében általában eleve adott modellel (például valamilyen eloszláscsaláddal) van dolgunk és annak ismeretlen paramétereire keresünk becsléseket, vagy végzünk rájuk vonatkozó hipotézisvizsgálatot, de ez nincs mindig így; előfordulhat ugyanis, hogy nincs semmi információnk az eloszláscsaládról, vagy csak annyi, hogy melyek jönnek szóba. Ekkor nekünk kell kiválasztanunk a megfelelő mértékcsaládot (likelihood) is, amit megtehetünk például a logaritmikus szkór segítségével. A logaritmikus szkór az egyik leggyakrabban használt szkór, mivel a prediktív eloszlástól csak a megfigyelésen keresztül függ¹ és a más területeken neki megfelelő fogalmak is igen elterjedtek: a becsléelméletben a ML-becslés, az információelméletben a Shannon-entrópia felel meg neki.

A logaritmikus szkórból kiszámolt átlagos jutalom alapján hozzuk meg döntésünket a modellek között, ezért ezek határeloszlását vizsgálom néhány ismert modell esetén. Ehhez *keresztkiértékelés*² módszert alkalmazok: felosztom a mintát két csoportra³, az egyikből becslöm az ismeretlen paraméter(ek)e)t⁴, míg a másodikkal úgy mond "validálom" az illesztett modellt: a logaritmikus szkór segítségével vizsgálom az illesztés minőségét. Az első csoportot általában úgy szokták megválasztani, hogy az elemszáma 70%-a legyen az eredeti minta elemszámának. Majd ezután két azonos várható értékű modellt hasonlítok össze, és vizsgálom, hogy mikor fogunk tévedni, hogyha a logaritmikus átlagszkór alapján döntünk közöttük. Most tehát különböző nevezetes modellek esetén vizsgálom meg hogyan alakul az átlagszkór határeloszlása.

Legyen X_1, \dots, X_n független minta valamilyen paraméteres eloszláscsaládból ismeretlen paraméterrel (vagy paraméterekkel). A mintát felosztom két részre, és az első mintából (*training data*) becslöm a paramétert, majd a második mintára (*test data*) felírt átlagszkór határeloszlását vizsgálom. Jelölje az így kapott két mintát $\underline{X}^{(1)}$, illetve $\underline{X}^{(2)}$. Először azt az esetet vizsgálom, mikor exponenciális eloszlást tételezek fel az adatokról:

¹Láttuk, hogy ez bizonyos értelemben ez fordítva is igaz.

²*Cross validation*

³*Training és test data*

⁴Maximum likelihood módszerrel, ugyanis az optimális szkór becslésnél láttuk, hogy ez felel meg a logaritmikus szkórnak.

4.1. Exponenciális modell

Exp(λ), ahol λ ismeretlen paraméter

Először tehát $\underline{X}^{(1)}$ -ből becslöm a λ paramétert maximum likelihood, hiszen ez a becslés felel meg a logaritmusos szórónak. Az így kapott becslés $\hat{\lambda} = (\overline{X^{(1)}})^{-1}$, ahol $\overline{X^{(1)}}$ jelöli az első minta átlagát. Ekkor a prediktív sűrűségfüggvényként a következő adódik:

$$p(x) = \hat{\lambda} e^{-\hat{\lambda}x}$$

Ezért a logaritmusos szóró felírva ebben az esetben így néz ki:

$$LS(p, x) = \log p(x) = \log \hat{\lambda} - \hat{\lambda}x$$

Most a megfigyeléseink az $\underline{X}^{(2)}$ minta elemei, így ezekre felírva a logaritmusos szórót, a következő átlagszórót kapjuk:

$$\overline{LS}(p, \underline{X}^{(2)}) := \frac{1}{n_2} \sum_{i=1}^{n_2} LS(p, X_i^{(2)}) = \log \hat{\lambda} - \frac{\hat{\lambda}}{\bar{\lambda}}$$

ahol $\tilde{\lambda}$ jelöli az $(\overline{X^{(2)}})^{-1}$ statisztikát. Ez éppen a λ paraméter második mintából származó ML-becslése. Tehát az $\overline{LS}(p, \underline{X}^{(2)})$ statisztika csak a két ML-becslés függvénye, így a határeloszlásához csak a ML-becslések aszimptotikus normalitását, és a következő, [1]-ben is megtalálható tételt használom fel:

Tétel: (A ML-becslés tulajdonságai)

(1) (RR) regularitási feltételek mellett az ML-becslés aszimptotikusan normális és optimális

$$\sqrt{n}(\vartheta_n - \vartheta) \longrightarrow \mathcal{N}(0, I(\vartheta)^{-1}) \quad (4.1)$$

ahol ϑ_n az ML-becslés, és $I(\vartheta)$ az egyelemű mintából számolt Fisher-információ.

(2) Ha

$$\sqrt{n}(t_n - \underline{\mu}) \longrightarrow \mathcal{N}_p(0, V) \quad (4.2)$$

és $\underline{f}=(f_1, \dots, f_q)$ differenciálható $\underline{\mu}$ -ben, akkor

$$\sqrt{n}(f(t_n) - f(\underline{\mu})) \longrightarrow \mathcal{N}_q(0, D^TVD) \quad (4.3)$$

ahol $D=(d_{ij}) = \left(\frac{\partial f_j}{\partial t_i}(\underline{\mu})\right)^5$

Alkalmazzuk most az ezeket az exponenciális eloszlás esetére. Mivel ekkor a Fisher-információ $I(\lambda) = 1/\lambda^2$, ezért $\hat{\lambda}$ -ra és $\tilde{\lambda}$ -ra a következő határeloszlások érvényesek:

$$\sqrt{n_1}(\hat{\lambda} - \lambda) \rightarrow^d \mathcal{N}(0, \lambda^2)$$

$$\sqrt{n_2}(\tilde{\lambda} - \lambda) \rightarrow^d \mathcal{N}(0, \lambda^2)$$

ahol n_1 , illetve n_2 jelöli $\underline{X}^{(1)}$ és $\underline{X}^{(2)}$ elemszámát. S mivel ezek független minták, az n_1 és n_2 mintaelemszámokra tett feltevések miatt ez a következőképpen is írható

$$\sqrt{n} \left(\begin{pmatrix} \hat{\lambda} \\ \tilde{\lambda} \end{pmatrix} - \begin{pmatrix} \lambda \\ \lambda \end{pmatrix} \right) \rightarrow^d \mathcal{N}_2(0, \Sigma), \quad \text{ahol } \Sigma = \begin{pmatrix} \frac{\lambda^2}{0,7} & 0 \\ 0 & \frac{\lambda^2}{0,3} \end{pmatrix}$$

⁵Vagyis egy aszimptotikusan normális statisztika függvénye is az (ha ez a függvény differenciálható az igazi paraméterértékben), és a kovarianciamátrix a deriváltmátrixszal transzformálódik.

A tételben szereplő f transzformációs függvény most a következő:

$$f(x, y) = \log x - \frac{x}{y}$$

Kiszámolva ennek a parciális deriváltjait (λ, λ) -ban kapjuk a D transzformációs mátrixot, ami ebben az esetben egy d vektor:

$$d = \begin{pmatrix} \partial_1 f \\ \partial_2 f \end{pmatrix} (\lambda, \lambda) = \begin{pmatrix} 0 \\ \frac{1}{\lambda} \end{pmatrix}$$

Így mikor az $f(\hat{\lambda}, \tilde{\lambda}) = \overline{LS}(p, \underline{X}^{(2)})$ -re írjuk fel az aszimptotikus normalitást a kovarianciamátrix a következőképpen transzformálódik: $d^T \Sigma d = \frac{10}{3}$.

Így tehát $\overline{LS}(p, \underline{X}^{(2)})$ -re a következő határeloszlás adódik:

$$\sqrt{n}(f(\hat{\lambda}, \tilde{\lambda}) - f(\lambda, \lambda)) = \sqrt{n} \left(\overline{LS}(p, \underline{X}^{(2)}) - \log \lambda + 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{10}{3} \right)$$

De mivel nem ismerjük λ valódi értékét, csak az első mintából származó becslését, ezért $f(\hat{\lambda}, \tilde{\lambda}) - f(\lambda, \lambda)$ helyett $f(\hat{\lambda}, \tilde{\lambda}) - f(\hat{\lambda}, \hat{\lambda})$ határeloszlását is kiszámolhatjuk:

$$\sqrt{n}(f(\hat{\lambda}, \tilde{\lambda}) - f(\hat{\lambda}, \hat{\lambda})) = \sqrt{n} \left(1 - \frac{\hat{\lambda}}{\tilde{\lambda}} \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{100}{21} \right)$$

4.2. Normális modell

$\mathcal{N}(\mu, \sigma^2)$, ahol mindkét paraméter ismeretlen

A következőben azt az esetet vizsgálom, mikor normális eloszlást tételezek fel az adatokról, melynek a paramétereit is becsülni kell és ekkor vizsgálom a logaritmikus átlagszkór határeloszlását. Ugyanúgy mint előbb, a mintát felosztjuk két részre: $\underline{X}^{(1)}$ -re és $\underline{X}^{(2)}$ -re, az elsőből becsülöm a paramétereket:

$$\hat{\mu} = \overline{X^{(1)}} \quad \text{és} \quad \hat{\sigma} = s_n^{(1)}$$

ahol $s_n^{(1)}$ jelöli az első minta tapasztalati szórását⁶

Tehát a prediktív sűrűségfüggvény

$$p(x) = \frac{1}{\sqrt{2\pi\hat{\sigma}}} \exp \left(-\frac{1}{2} \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right)^2 \right)$$

A logaritmikus szkórt az $\underline{X}^{(2)}$ minta elemeire felírva, a következő átlagszkórt kapom:

$$\overline{LS}(p, \underline{X}^{(2)}) = -\frac{1}{2} \log 2\pi - \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} \left(\tilde{\sigma}^2 + (\hat{\mu} - \tilde{\mu})^2 \right)$$

Itt $\tilde{\mu}$ és $\tilde{\sigma}$ jelöli a második mintából származó ML-becsléseket. Az átlagszkór most is csak a ML-becslések függvénye és ezért megint elővehető a ML-becslések konzisztenciája és aszimptotikus normalitása:

$$\overline{LS}(p, \underline{X}^{(2)}) \xrightarrow{d} -\frac{1}{2} \log 2\pi - \log \sigma - \frac{1}{2}$$

⁶ $s_n^{(1)} = \left(\frac{1}{n_2} \sum_{i=1}^{n_2} (X_i^{(2)} - \overline{X^{(2)}})^2 \right)^{1/2}$ (ez σ ML-becslése lesz)

A határeloszláshoz ki kell számolni a Fisher-információt: eltolás- és skála-paraméteres családban ez nem függ az eltolásparamétertől (μ), és a skála-paramétertől (σ) is csak lineárisan

$$I(\mu, \sigma) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Fontos, hogy nem σ^2 -ben, hanem σ -ban kaptuk meg a Fisher-információt. Persze átparaméterezéssel megkaphatnánk σ^2 -ben is, de szükségtelen, mivel úgyis σ ML-becslését írtuk fel korábban.

Így a $\hat{\mu}, \tilde{\mu}, \hat{\sigma}, \tilde{\sigma}$ ML-becslések aszimptotikus normalitását együttesen felírva és felhasználva, hogy a két minta is és a $\hat{\mu}, \hat{\sigma}$ (ill. a $\tilde{\mu}, \tilde{\sigma}$) becslések is függetlenek egymástól⁷, a következő adódik:

$$\sqrt{n} \left[\begin{pmatrix} \hat{\mu} \\ \tilde{\mu} \\ \hat{\sigma} \\ \tilde{\sigma} \end{pmatrix} - \begin{pmatrix} \mu \\ \mu \\ \sigma \\ \sigma \end{pmatrix} \right] \rightarrow^d \mathcal{N}_4(0, \Sigma), \quad \text{ahol} \quad \Sigma = \begin{pmatrix} \frac{\sigma^2}{0,7} & 0 & 0 & 0 \\ 0 & \frac{\sigma^2}{0,3} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{1,4} & 0 \\ 0 & 0 & 0 & \frac{\sigma^2}{0,6} \end{pmatrix}$$

Az $\overline{LS}(p, \underline{X}^{(2)})$ statisztikát előállító transzformációs függvény ebben az esetben:

$$f(x_1, x_2, x_3, x_4) = -\frac{1}{2} \log 2\pi - \log x_3 - \frac{1}{2} \left(\frac{x_4}{x_3} \right)^2 - \frac{1}{2} \left(\frac{x_1 - x_2}{x_3} \right)^2$$

Kiszámolva a parciális deriváltakat $(\mu, \mu, \sigma, \sigma)$ -ban a transzformáló vektor ekkor:

$$d = \text{grad}f(\mu, \mu, \sigma, \sigma) = \left(0, 0, 0, -\frac{1}{\sigma} \right)$$

Így az átlagszkórra a következő határeloszlás adódik

$$\sqrt{n} \left(\overline{LS}(p, \underline{X}^{(2)}) + \frac{1}{2} \log 2\pi + \log \sigma + \frac{1}{2} \right) \rightarrow^d \mathcal{N}(0, 10/6)$$

ugyanis a transzformált kovarianciamátrix $d^T \Sigma d = \frac{10}{6}$.

Megintcsak a σ paraméter valódi értéke nem ismert, ezért $\sqrt{n}(f(\hat{\mu}, \tilde{\mu}, \hat{\sigma}, \tilde{\sigma}) - f(\mu, \mu, \sigma, \sigma))$ helyett a $\sqrt{n}(f(\hat{\mu}, \tilde{\mu}, \hat{\sigma}, \tilde{\sigma}) - f(\hat{\mu}, \hat{\mu}, \hat{\sigma}, \hat{\sigma}))$ határeloszlását kiszámolva:

$$\sqrt{n} \left(-\frac{1}{2\hat{\sigma}^2} (\hat{\sigma}^2 + (\hat{\mu} - \tilde{\mu})^2 - \hat{\sigma}^2) \right) \rightarrow^d \mathcal{N} \left(0, \frac{50}{21} \right)$$

Megfigyelhetjük, hogy mindkét esetben feleakkora aszimptotikus szórásnégyzetet tapasztalunk mint az exponenciális modellnél. Ez persze nem annyira meglepő mivel a normális modellben két paramétert becsültünk (míg az exponenciálisnál csak egyet); és így természetes, hogy aszimptotikusan jobb eredményt kapunk az utóbbi esetben, hiszen ekkor jobban be tudjuk állítani a prediktív eloszlást. Persze a normális eloszlás az egész számegyenesen pozitív sűrűségfüggvénnyel rendelkezik, míg az exponenciális eloszlás csak a pozitív félegyenesre van koncentrálna, így nem igazán lehet a két illesztést összehasonlítani. De ha elég nagy számú mintát veszünk, akkor a kapott megfigyelésekből már tudunk az eloszlás tartójára következtetni. Most két olyan modellt vizsgálok, amik szintén a nemnegatív félegyenesre vannak koncentrálna. Az így kapott eredmények inkább összevethetőek az exponenciális eloszlásnál kapottakkal. Az először a lognormális modellt vizsgálom.

⁷Fisher-Bartlett tétel [1]

4.3. Lognormális modell

LogNorm(μ, σ^2), ahol mindkét paraméter ismeretlen

Vegyünk mintát az eloszlásból és osszuk fel két részre: $\underline{X}^{(1)}$ -re és $\underline{X}^{(2)}$ -re. Mivel a kapott adatokról lognormális eloszlást feltételezünk, ezért mikor az első csoportból becslöm a paramétereket (maximum likelihood módszerrel), a mintaelemek logaritmusainak normalitását kihasználva könnyen adódik:

$$\hat{\mu} = \overline{\log X^{(1)}} \text{ az } \underline{X}^{(1)}\text{-beli mintaelemek logaritmusainak átlaga}$$

$$\hat{\sigma} \text{ pedig a mintaelemek logaritmusainak tapasztalati szórása}$$

Ekkor a prediktív sűrűségfüggvényünk a következő lesz:

$$p(x) = \frac{1}{\hat{\sigma}x\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log x - \hat{\mu}}{\hat{\sigma}}\right)^2\right)$$

A logaritmikus átlagszkórt kiszámolva a következő adódik:

$$\overline{LS}(p, \underline{X}^{(2)}) = -\tilde{\mu} - \frac{1}{2} \log 2\pi - \log \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} (\tilde{\sigma}^2 + (\hat{\mu} - \tilde{\mu})^2)$$

Itt szintén $\tilde{\mu}$ és $\tilde{\sigma}$ jelöli a második mintából származó ML-becsléseket. Most a becslések más statisztikákat jelölnek, amiknek viszont az aszimptotikus viselkedései megegyeznek a normális eloszlásnál tapasztaltakkal. Így a Fisher-információ sem változik: $I(\mu, \sigma) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

Ezért ezekre is igaz a következő:

$$\sqrt{n} \left[\begin{pmatrix} \hat{\mu} \\ \tilde{\mu} \\ \hat{\sigma} \\ \tilde{\sigma} \end{pmatrix} - \begin{pmatrix} \mu \\ \mu \\ \sigma \\ \sigma \end{pmatrix} \right] \xrightarrow{d} \mathcal{N}_4(0, \Sigma), \quad \text{ahol } \Sigma = \begin{pmatrix} \frac{\sigma^2}{0,7} & 0 & 0 & 0 \\ 0 & \frac{\sigma^2}{0,3} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{1,4} & 0 \\ 0 & 0 & 0 & \frac{\sigma^2}{0,6} \end{pmatrix}$$

Az egyetlen különbség a normális modellhez képest a transzformációs függvényünkben megjelenő x_2 új tag:

$$f(x_1, x_2, x_3, x_4) = -\frac{1}{2} \log 2\pi - x_2 - \log x_3 - \frac{1}{2} \left(\frac{x_4}{x_3}\right)^2 - \frac{1}{2} \left(\frac{x_1 - x_2}{x_3}\right)^2$$

Ennek gradiense

$$d^T = (\partial_1 f, \partial_2 f, \partial_3 f, \partial_4 f)(\mu, \mu, \sigma, \sigma) = (0, 1, 0, -\frac{1}{\sigma})$$

és így az új kovarianciamátrix $d^T \Sigma d = \frac{\sigma^2}{0,3} + 10/6$ lesz.

Tehát a logaritmikus átlagszkó aszimptotikusan normális a következő kovarianciamátrixsal:

$$\sqrt{n} \left(\overline{LS}(p, \underline{X}^{(2)}) + \mu + \frac{1}{2} \log 2\pi + \log \sigma + \frac{1}{2} \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{5}{3} (2\sigma^2 + 1) \right)$$

S mivel a $2\sigma^2 + 1$ σ folytonos függvénye, ezért ez a következőképpen is írható:

$$\sqrt{n} \left(\frac{\overline{LS}(p, \underline{X}^{(2)}) + \mu + \frac{1}{2} \log 2\pi + \log \sigma + \frac{1}{2}}{(2\hat{\sigma}^2 + 1)} \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{5}{3} \right)$$

4.4. Gamma modell

$\Gamma_{\alpha,\lambda}$, ahol mindkét paraméter ismeretlen

Vegyünk egy \underline{X} mintát és szokás szerint osszuk fel két részre: $\underline{X}^{(1)}$ -re és $\underline{X}^{(2)}$ -re. Most az α paraméter maximum likelihood becslésének kiszámítás kicsit nehézkes lenne, ezért momentum módszerrel becsljük a paramétereket, melyre szintén könnyen bizonyítható aszimptotikus normalitás:

Tétel [1]:

Legyenek a statisztikáink a következő "átlag" alakúak:

$T_1 = \frac{1}{n} \sum_{i=1}^n g_1(x_i), \dots, T_k = \frac{1}{n} \sum_{i=1}^n g_k(x_i)$. Ekkor a $T = (T_1, \dots, T_k)$ k-dimenziós statisztika aszimptotikusan normális:

$$\sqrt{n} \left[\begin{pmatrix} T_1 \\ \vdots \\ T_k \end{pmatrix} - \begin{pmatrix} E(g_1(X)) \\ \vdots \\ E(g_k(X)) \end{pmatrix} \right] \longrightarrow^d \mathcal{N}_k(0, \Sigma)$$

ahol $\Sigma = (\Sigma_{ij}) = (\text{cov}(g_i(X), g_j(X)))$

Most 2 ismeretlen paraméterünk van, ezért az első kettő momentumra írjuk fel a tételt:

$$\sqrt{n} \left[\begin{pmatrix} \bar{X} \\ \bar{X}^2 \end{pmatrix} - \begin{pmatrix} E(X) \\ E(X^2) \end{pmatrix} \right] \longrightarrow^d \mathcal{N}_2(0, \Sigma)$$

ahol $\Sigma = \begin{pmatrix} D^2(X) & \text{cov}(X, X^2) \\ \text{cov}(X^2, X) & D^2(X^2) \end{pmatrix}$

A paraméterekre a momentum módszerrel a következő becsléseket kapjuk:

$$\hat{\alpha} = \left(\frac{\bar{X}^{(1)}}{s_n^{(1)}} \right)^2, \quad \hat{\lambda} = \frac{\bar{X}^{(1)}}{(s_n^{(1)})^2}$$

ahol $\bar{X}^{(1)}$, ill. $s_n^{(1)}$ jelöli az első minta tapasztalati közepét és szórását.

A prediktív sűrűségfüggvény tehát a következő:

$$p(x) = \frac{\hat{\lambda}^{\hat{\alpha}}}{\Gamma(\hat{\alpha})} x^{\hat{\alpha}-1} e^{-\hat{\lambda}x}$$

Most

$$\bar{X}^{(1)} = \frac{\hat{\alpha}}{\hat{\lambda}}\text{-ra és } \overline{(X^{(1)})^2} = \frac{\hat{\alpha}(\hat{\alpha}+1)}{\hat{\lambda}^2}\text{-re}$$

tudjuk az aszimptotikus normalitást

$$\sqrt{n} \left[\begin{pmatrix} \frac{\hat{\alpha}}{\hat{\lambda}} \\ \frac{\hat{\alpha}(\hat{\alpha}+1)}{\hat{\lambda}^2} \end{pmatrix} - \begin{pmatrix} \frac{\alpha}{\lambda} \\ \frac{\alpha(\alpha+1)}{\lambda^2} \end{pmatrix} \right] \longrightarrow^d \mathcal{N}(0, \Sigma)$$

ahol $\Sigma = \begin{pmatrix} \frac{\alpha}{\lambda^2} & \frac{2\alpha(\alpha+1)}{\lambda^3} \\ \frac{2\alpha(\alpha+1)}{\lambda^3} & \frac{2\alpha(\alpha+1)(2\alpha+3)}{\lambda^4} \end{pmatrix}$

Következő lépésben felírjuk a logaritmusos átlagszkórt az $X^{(2)}$ elemeiben:

$$\overline{LS(p, \underline{X}^{(2)})} = \hat{\alpha} \log \hat{\lambda} - \log \Gamma(\hat{\alpha}) + (\hat{\alpha} - 1) \overline{\log X^{(2)}} - \hat{\lambda} \overline{X^{(2)}}$$

Ki lehetne fejezni ezt a fenti "átlag" alakú statisztikák függvényeként is, de csak bonyolítaná a képletet. A $\overline{\log X^{(2)}}$ -re és $\overline{X^{(2)}}$ -re a tétel szintén biztosítja az aszimptotikus normalitást:

$$\sqrt{n} \left[\begin{pmatrix} \overline{\log X^{(2)}} \\ \overline{X^{(2)}} \end{pmatrix} - \begin{pmatrix} \Psi(\alpha) - \log \lambda \\ \frac{\alpha}{\lambda} \end{pmatrix} \right] \rightarrow^d \mathcal{N}(0, \Sigma)$$

$$\text{ahol}^8 \Sigma = \begin{pmatrix} D^2(\log X) & \text{cov}(X, \log X) \\ \text{cov}(X, \log X) & D^2(X) \end{pmatrix} = \begin{pmatrix} \Psi(1, \alpha) & \frac{1}{\lambda} \\ \frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{pmatrix}$$

Így ezekre és az $X^{(1)}$ -beli statisztikákra együttesen is felírva az aszimptotikus normalitást: a következő kovarianciamátrixot kapjuk a két minta függetlensége miatt:

$$\Sigma = \begin{pmatrix} \frac{\alpha}{\lambda^2} & \frac{2\alpha(\alpha+1)}{\lambda^3} & 0 & 0 \\ \frac{2\alpha(\alpha+1)}{\lambda^3} & \frac{2\alpha(\alpha+1)(2\alpha+3)}{\lambda^4} & 0 & 0 \\ 0 & 0 & \Psi(1, \alpha) & \frac{1}{\lambda} \\ 0 & 0 & \frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{pmatrix}$$

Most a transzformációs függvényünk amivel az átlagszkór statisztikát kapjuk:

$$h(x_1, x_2, x_3, x_4) = -\frac{x_1^2}{x_2 - x_1^2} \log \frac{x_1}{x_2 - x_1^2} + \log \Gamma \left(\frac{x_1^2}{x_2 - x_1^2} \right) - \left(\frac{x_1^2}{x_2 - x_1^2} - 1 \right) x_3 + \frac{x_1}{x_2 - x_1^2} x_4$$

Kiszámolva a parciális deriváltjait a $(\frac{\alpha}{\lambda}, \frac{\alpha(\alpha+1)}{\lambda^2}, \Psi(\alpha) - \log \lambda, \frac{\alpha}{\lambda})$ pontban a következő transzformáló vektort kapjuk:

$$d^T = ((2\alpha + 1)(\alpha/\lambda - \lambda), 0, 1 - \alpha, \lambda)$$

Ezután fel lehet írni a transzformált kovarianciamátrixot is ezek segítségével. Ehelyett inkább vizsgáljunk meg egy "másik" gamma modellt.

$\Gamma_{\alpha, \alpha\lambda}$, ahol az α ismert, míg a λ ismeretlen paraméter

A következőben a gamma modell egy speciális esetét vizsgálom, amely majd a következőkben még szerephez jut. A paramétereit úgy választom, hogy a várható értéke λ legyen.

Vegyük az \underline{X} mintát, és osszuk fel $\underline{X}^{(1)}$ -re és $\underline{X}^{(2)}$ -re. Az elsőből becsüljük az ismeretlen λ paramétert: $\hat{\lambda} = \frac{1}{\underline{X}^{(1)}}$. A $p(x)$ prediktív sűrűségfüggvény most a következő:

$$p(x) = \frac{(\alpha\hat{\lambda})^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\alpha\hat{\lambda}x}$$

A logaritmikusan átlagszkór pedig

$$\overline{LS}(p, \underline{X}^{(2)}) = \alpha \log(\alpha\hat{\lambda}) - \log \Gamma(\alpha) + (\alpha - 1) \overline{\log X^{(2)}} - \alpha \frac{\hat{\lambda}}{\lambda}$$

⁸ $\Psi(\alpha) - \log \lambda = E \log X$, $\Psi(\alpha) = \partial_\alpha \log \Gamma(\alpha) = \frac{\partial_\alpha \Gamma(\alpha)}{\Gamma(\alpha)}$, $\Psi(1, \alpha) = \partial_\alpha \Psi(\alpha)$

Az ebben szereplő statisztikákra tudjuk az aszimptotikus normalitást:

$$\sqrt{n} \left[\begin{pmatrix} \frac{\overline{X^{(1)}}}{\overline{X^{(2)}}} \\ \frac{1}{\log X^{(2)}} \end{pmatrix} - \begin{pmatrix} \frac{1}{\lambda} \\ \Psi(\alpha) - \log(\alpha\lambda) \end{pmatrix} \right] \rightarrow^d \mathcal{N}(0, \Sigma)$$

ahol

$$\Sigma = \begin{pmatrix} \frac{1}{0,7 \cdot \alpha \lambda^2} & 0 & 0 \\ 0 & \frac{1}{0,3 \cdot \alpha \lambda^2} & \frac{1}{0,3 \cdot \alpha \lambda} \\ 0 & \frac{1}{0,3 \cdot \alpha \lambda} & \frac{\Psi(1, \alpha)}{0,3} \end{pmatrix}$$

Ez esetben a következő a transzformációs függvényünk:

$$f(x, y, z) = -\alpha \log \alpha + \alpha \log x + \log \Gamma(\alpha) - (\alpha - 1)z + \alpha \frac{y}{x}$$

Ennek a parciális deriváltjai a $(\frac{1}{\lambda}, \frac{1}{\lambda}, \Psi(\alpha) - \log(\alpha\lambda))$ pontban:

$$\text{grad} f \left(\frac{1}{\lambda}, \frac{1}{\lambda}, \Psi(\alpha) - \log(\alpha\lambda) \right) = (0, \alpha\lambda, 1 - \alpha)$$

Így az átlagszkórra a következő határeloszlás adódik:

$$\begin{aligned} & \sqrt{n} \left(\overline{LS}(p, X^{(2)}) - \alpha \log(\alpha\lambda) + \log \Gamma(\alpha) - (\alpha - 1)\Psi(\alpha) + \alpha \right) = \\ & \sqrt{n} \left(f \left(\begin{pmatrix} \frac{\overline{X^{(1)}}}{\overline{X^{(2)}}} \\ \frac{1}{\log X^{(2)}} \end{pmatrix} \right) - f \left(\begin{pmatrix} \frac{1}{\lambda} \\ \Psi(\alpha) - \log(\alpha\lambda) \end{pmatrix} \right) \right) \rightarrow^d \mathcal{N}(0, \sigma^2) \end{aligned}$$

$$\text{ahol } \sigma^2 = \frac{1}{0,3} (2 - \alpha + (1 - \alpha)^2 \Psi(1, \alpha))$$

Megintcsak nem ismerjük λ valódi értékét, csak a $\hat{\lambda}$ becslését, ezért erre is felírható határeloszlás:

$$\sqrt{n} \left(\overline{LS}(p, X^{(2)}) - \alpha \log(\alpha\hat{\lambda}) + \log \Gamma(\alpha) - (\alpha - 1)\Psi(\alpha) + \alpha \right) \rightarrow^d \mathcal{N}(0, \sigma^{*2})$$

$$\text{ahol } \sigma^{*2} = \frac{1}{0,7 \cdot \alpha} + \frac{1}{0,3} (2 - \alpha + (1 - \alpha)^2 \Psi(1, \alpha)) \quad \square$$

A következőben azt az esetet vizsgálom, amikor van elképzelésünk vagy információnk a szóba jöhető modellekről, és ezek közül szeretnénk kiválasztani a legjobbat. Legegyszerűbb eset, amikor két modell közül szeretnénk kiválasztani a helyeset. Az egyik legyen az exponenciális, a másik pedig az előbb már tárgyalt gamma modell, ahol ez utóbbi α rendjét ismertnek tekintem, a másik paramétert pedig úgy választom meg, hogy a várható érték a két modellben megegyezzen:

Exp(λ) vs $\Gamma_{\alpha, \alpha\lambda}$ (ahol α ismert, λ ismeretlen paraméter)

Mindkét modellben tehát λ a várható érték, de ez ismeretlen, így becsülni kell. Vegyünk tehát mintát az ismeretlen eloszlásból és osszuk fel szokásosan két részre. Ez elsőből becsüljük a λ paramétert a mintaátlag reciprokával. Így a két modellnek megfelelő prediktív sűrűségfüggvények:

$$\hat{f}(x) = \hat{\lambda} \exp(-\hat{\lambda}x)$$

és

$$\hat{g}(x) = \frac{(\alpha \hat{\lambda})^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\alpha \hat{\lambda} x)$$

Ekkor vizsgálhatjuk a tévedés valószínűségét: tévedhetünk úgy is, hogy ha az exponenciális modell mellett döntünk a logaritmikus átlagszór alapján, mikor valójában a gamma modell a helyes és ez történhet fordítva is. Tehát afféle első-, ill. másodrendű hibákat keresünk. Először nézzük meg, hogy mikor tévedünk az exponenciális modell kárára, azaz mely esetekben választjuk a gamma modellt, amikor az exponenciális lenne a helyes döntés. Ez akkor történik meg, ha a logaritmikus átlagszór a gamma modell esetén a nagyobb. Vagyis akkor tévedünk, hogyha:

$$\overline{LS}(\hat{f}, \underline{X}^{(2)}) \leq \overline{LS}(\hat{g}, \underline{X}^{(2)}) \quad (4.4)$$

Ezt részletesen kiírva és a $\tilde{\lambda} = \frac{1}{\overline{X}^{(2)}}$, illetve $\overline{\log X^{(2)}}$ jelöléseket bevezetve:

$$\log \hat{\lambda} - \frac{\hat{\lambda}}{\tilde{\lambda}} \leq \alpha \log \alpha + \alpha \log \hat{\lambda} - \log \Gamma(\alpha) + (\alpha - 1) \overline{\log X^{(2)}} - \alpha \frac{\hat{\lambda}}{\tilde{\lambda}} \quad (4.5)$$

Mivel az α ismert, ezért mindkét oldal exponensét vesszük és az α -t tartalmazó tagokat az egyik oldalra gyűjtjük⁹

$$c(\alpha) := \left(\frac{\alpha^\alpha}{\Gamma(\alpha)} \right) \geq \left(\exp \left(\frac{\hat{\lambda}}{\tilde{\lambda}} - \overline{\log X^{(2)}} \right) \frac{1}{\tilde{\lambda}} \right)^{\alpha-1} \quad (4.6)$$

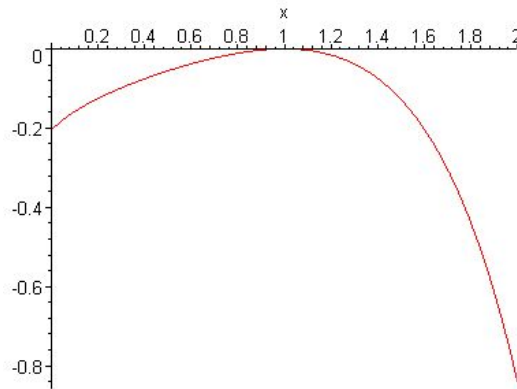
Ha most növeljük a minta elemszámát, akkor a bal oldal konvergens, $\exp(1 - E(\log X)) \frac{1}{\tilde{\lambda}}$ -hoz tart, ugyanis $\log X^{(2)}$ a nagy számok erős törvénye szerint a megfelelő várható értékhez tart, most az exponenciális eloszlást tételeztük fel helyesnek, így a várható értéket is e szerint kell képezni:

$$\overline{\log X^{(2)}} \rightarrow E_{\exp}(\log X) = \int_0^\infty \log x \lambda e^{-\lambda x} dx = -\gamma - \log \lambda$$

ahol γ az Euler-konstans. A bal oldal nem függ az n mintaelemszámtól, így a következő határátmenetet kapjuk:

$$c(\alpha) \geq \exp((1 + \gamma)(\alpha - 1))$$

Tehát egyik oldal se függ λ -tól, így aszimptotikusan az α paramétertől függ, hogy mikor tévedünk. Ábrázolva a két oldal különbségét láthatjuk, hogy az nem pozitív, így aszimptoti-



4.1. ábra. $c(\alpha) - \exp((1 + \gamma)(\alpha - 1))$

⁹Itt vigyázni kell mert amikor $\alpha < 1$. gyököt vonunk mindkét oldalból, akkor $\alpha < 1$ esetén megváltozik az egyenlőtlenség.

kusan mindig a jó modellt fogjuk kiválasztani, ha a logaritmusos átlagszór alapján döntünk.

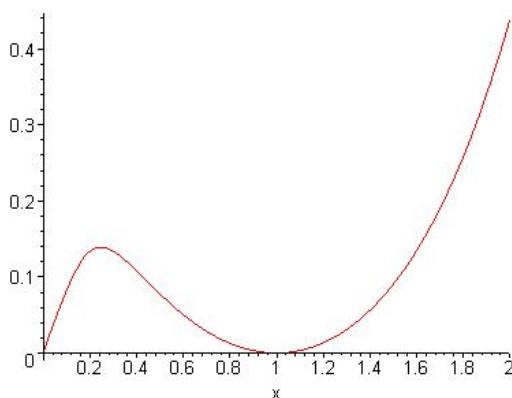
Nézzük meg most a másik esetet, mikor a gamma modell a helyes, de az exponenciális válasszuk a logaritmusos átlagszór alapján. Az eddigi jelöléseket használva minden hasonlóan történik, csak (1)-ben az egyenlőtlenség épp fordítva áll és (3)-ban a jobb oldalon levő statisztika határértéke most más lesz, ugyanis most $\overline{\log X^{(2)}}$ várható értékét a (valódi) gamma eloszlás szerint kell venni:

$$\overline{\log X^{(2)}} \rightarrow E_{\Gamma}(\log X) = \int_0^{\infty} \log x \frac{(\alpha\lambda)^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha\lambda x} dx = \Psi(\alpha) - \log(\alpha\lambda)^{10}$$

Ekkor a következőt kapjuk határátmenetként:

$$c(\alpha) \leq \left(\alpha e^{(1-\Psi(\alpha))} \right)^{\alpha-1}$$

Megintcsak kiestek a λ -t tartalmazó tagok, így aszimptotikusan csak α -tól függ, hogy mikor tévedünk. Ismét ábrázolva a két oldal különbségét, ez nemnegatív, így itt is 1-hez tartó valószínűséggel a jó modellt fogjuk kiválasztani, ha a logaritmusos átlagszór alapján hozzuk meg döntésünket.



4.2. ábra. $c(\alpha) = \left(\alpha e^{(1-\Psi(\alpha))} \right)^{\alpha-1}$

Most azt az esetet vizsgálom amikor mindegyik mintaelemhez tartozik egy t_i megfigyelési súly (például mennyi ideig figyeltük meg az elemet), melyet ismertnek tekintünk és az előzőeket próbálom általánosítani erre az esetre. A t_i -vel kapcsolatban a következő enyhe megszorítással élek: tegyük fel, hogy a $\frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)}$ összeg határértéke létezik és véges.¹¹

Tehát az $\mathbf{Exp}\left(\frac{\lambda}{t_i}\right)$ és a $\Gamma\left(\alpha, \frac{\alpha\lambda}{t_i}\right)$ modellek közül szeretnénk kiválasztani a valódit, most is mindkét esetben λt_i a várható érték. Ebben az esetben a megfigyeléseink $X_i = t_i Y_i$ alakúak, ahol Y_i -k független, azonos eloszlású változók az egyik eloszlásból. Becsülnünk kell az ismeretlen λ paramétert, ehhez vegyünk ismét mintát az eloszlásból és osszuk fel két részre: $\underline{X}^{(1)}$ -re és $\underline{X}^{(2)}$ -re. A ML-becsléshez írjuk fel az első minta együttes sűrűségfüggvényét. Az i . mintaelem sűrűségfüggvénye az exponenciális esetben

$$f_{X_i}(x) = \frac{\lambda}{t_i} \exp\left(-\frac{\lambda}{t_i} x\right)$$

Az egész minta együttes sűrűségfüggvénye ekkor:

$$f_{X^{(1)}}(\underline{x}) = \prod_{i=1}^{n_1} f_{X_i}(x_i) = \frac{\lambda^{n_1}}{\prod_{i=1}^{n_1} t_i^{(1)}} \exp\left(-\lambda \sum_{i=1}^{n_1} \frac{x_i^{(1)}}{t_i^{(1)}}\right)$$

¹⁰ Ahol $\Psi(\alpha) = \partial_{\alpha} \log \Gamma(\alpha) = \frac{\partial_{\alpha} \Gamma(\alpha)}{\Gamma(\alpha)}$, a Γ -függvény logaritmusos deriváltja.

¹¹ Az előző esetben, mikor mindegyik t_i eggyel volt egyenlő, ez azonosan 0.

A loglikelihood-függvényt ennek logaritmusaként kapjuk:

$$l(\lambda) = n_1 \log \lambda - \sum_{i=1}^{n_1} \log t_i^{(1)} - \lambda \sum_{i=1}^{n_1} (x_i^{(1)}/t_i^{(1)})$$

Ezt λ szerint lederiválva, adódik a ML-becsolás:

$$\hat{\lambda} = \frac{n_1}{\sum_{i=1}^{n_1} (x_i^{(1)}/t_i^{(1)})}$$

Nézzük meg, hogy ugyanezt kapjuk-e a Γ modell feltételezve helyesnek. Ekkor a megfigyelések ismét $X_i = t_i Y_i$ alakúak, ahol Y_i -k független, azonos eloszlású $\Gamma_{\alpha, \alpha\lambda}$ változók. Az i . elem sűrűségfüggvénye ebben az esetben

$$g_{X_i}(x) = \left(\frac{\alpha\lambda}{t_i}\right)^\alpha \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{\alpha\lambda}{t_i} x\right)$$

A minta együttes sűrűségfüggvénye

$$g_{X^{(1)}}(\underline{x}) = \prod_{i=1}^{n_1} g_{X_i}(x_i) = \frac{(\alpha\lambda)^{\alpha n_1}}{(\prod_{i=1}^{n_1} t_i^{(1)})^\alpha} \frac{\prod_{i=1}^{n_1} x_i^{\alpha-1}}{\Gamma(\alpha)^{n_1}} \exp\left(-\alpha\lambda \sum_{i=1}^{n_1} \frac{x_i^{(1)}}{t_i^{(1)}}\right)$$

Logaritmust véve majd λ szerint lederiválva ugyanazt a ML-becsolást kapjuk:

$$\hat{\lambda} = \frac{n_1}{\sum_{i=1}^{n_1} (x_i^{(1)}/t_i^{(1)})}$$

Ezért jelölje

$$\hat{f}(x) := \hat{\lambda} \exp(-\hat{\lambda}x) \quad \text{és} \quad \hat{g}(x) := \frac{\alpha\hat{\lambda}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\alpha\hat{\lambda}x)$$

a $\hat{\lambda}$ -paraméterű exponenciális és a megfelelő Γ sűrűségfüggvényt.

Ekkor az $X_j^{(2)}$ mintaelem prediktív sűrűségfüggvénye $1/t_j^{(2)} \cdot \hat{f}(x/t_j^{(2)})$ lesz.

A logaritmikus szkór $X_j^{(2)}$ -ben felírva a következő alakú lesz:

$$LS(\hat{f}, X_j^{(2)}) = \log\left(1/t_j^{(2)} \hat{f}(x/t_j^{(2)})\right) = \log \hat{\lambda} - \log t_j^{(2)} - \hat{\lambda} (x_j^{(2)}/t_j^{(2)})$$

Most a második mintából mindegyik mintaelemnek különböző a prediktív sűrűségfüggvénye¹²; az átlagszkór ekkor így alakul:

$$\overline{LS}(\hat{f}, \underline{X}^{(2)}) = \frac{1}{n_2} \sum_{j=1}^{n_2} LS(\hat{f}, X_j^{(2)}) = \log \hat{\lambda} - \frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)} - \frac{\hat{\lambda}}{\hat{\lambda}}$$

Legyen $K := \lim_{n_2} \frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)}$, ekkor

$$\overline{LS}(\hat{f}, \underline{X}^{(2)}) \longrightarrow \log \lambda - K - 1$$

Most nézzük meg a Γ modell esetén mit kapunk az átlagszkórra. Ekkor az $X_j^{(2)}$ mintaelem prediktív sűrűségfüggvénye $1/t_j^{(2)} \cdot \hat{g}(x/t_j^{(2)})$ lesz, hasonlóan az exponenciális eloszláshoz, most is mindegyik sűrűségfüggvény $t_j^{(2)}$ -vel van skálatranszformálva.

A logaritmikus szkór $X_j^{(2)}$ -ben:

$$LS(\hat{g}, X_j^{(2)}) = \log\left(1/t_j^{(2)} \hat{g}(x/t_j^{(2)})\right) =$$

¹²Épp $t_j^{(2)}$ -vel van skálatranszformálva \hat{f} .

$$\alpha \log \alpha + \alpha \log \hat{\lambda} - \log t_j^{(2)} - \log \Gamma(\alpha) + (\alpha - 1) \log(X_j^{(2)}/t_j^{(2)}) - \alpha \hat{\lambda}(X_j^{(2)}/t_j^{(2)})$$

Az átlagszkór a következő:

$$\begin{aligned} \overline{LS}(\hat{g}, \underline{X}^{(2)}) &= \frac{1}{n_2} \sum_{j=1}^{n_2} LS(\hat{g}, X_j^{(2)}) = \alpha \log \alpha + \alpha \log \hat{\lambda} - \log \Gamma(\alpha) - \alpha \frac{\hat{\lambda}}{\lambda} - \\ &\quad \frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)} + (\alpha - 1) \frac{1}{n_2} \sum_{j=1}^{n_2} \log(X_j^{(2)}/t_j^{(2)}) \end{aligned}$$

Most, ha az n mintaelemszámmal végtelenhez tartunk, akkor

$$\overline{LS}(\hat{g}, \underline{X}^{(2)}) \longrightarrow -\alpha \log \alpha - \alpha \log \lambda + \log \Gamma(\alpha) + \alpha + K - (\alpha - 1)E(\log Y)$$

Ugyanis $X_j^{(2)}/t_j^{(2)} = Y_j^{(2)}$ független, azonos eloszlásúak és $\frac{1}{n_2} \sum_{j=1}^{n_2} \log(Y_j^{(2)}) \longrightarrow E(\log Y)$, ahol Y eloszlása az $Y_j^{(2)}$ mintaelemek valódi eloszlásával egyezik meg.

Most is azt az esetet vizsgálom először, mikor az exponenciális modell a helyes és tévedünk mikor Γ -előrejelzést adunk. Ezért az $Y_j^{(2)}$ -k valójában λ -paraméterű exponenciális eloszlásúak, ezért a várható értéket is e szerint kell képezzük és így: $E(\log Y) = -\gamma - \log \lambda$. Megintcsak akkor adjuk a helytelen Γ -előrejelzést, ha az átlagszkór ebben az esetben lesz kisebb.

$$\begin{aligned} \overline{LS}(\hat{f}, \underline{X}^{(2)}) &= -\log \hat{\lambda} + \frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)} + \frac{\hat{\lambda}}{\lambda} \geq \overline{LS}(\hat{g}, \underline{X}^{(2)}) = \\ &= -\alpha \log \alpha - \alpha \log \hat{\lambda} + \log \Gamma(\alpha) + \alpha \frac{\hat{\lambda}}{\lambda} + \frac{1}{n_2} \sum_{j=1}^{n_2} \log t_j^{(2)} - (\alpha - 1) \frac{1}{n_2} \sum_{j=1}^{n_2} \log(X_j^{(2)}/t_j^{(2)}) \end{aligned}$$

Ha a mintaelemszámmal végtelenhez tartunk és aszimptotikusan vizsgáljuk a két oldalt, akkor a következőt kapjuk:

$$-\log \lambda + K + 1 \geq -\alpha \log \alpha - \alpha \log \lambda + \log \Gamma(\alpha) + \alpha + K + (\alpha - 1)(\gamma + \log \lambda)$$

A λ -t tartalmazó tagok ismét kiejtik egymást, s mindkét oldalt csak egy új tag (K) jelent meg az előző esethez képest, tehát ismét elmondható, hogyha a logaritmikus átlagszkór alapján hozzuk meg döntésünk, akkor 1-hez tartó valószínűséggel a helyes modellt fogjuk kiválasztani.

Most nézzük meg ezt általánosan is. Legyen Q a valódi háttéreloszlás és legyen P az előrejelzés, és vizsgáljuk meg mikor tévedünk. Ez pontosan akkor következik be, ha a minta alapján számolt átlagszkórokra a következő teljesül:

$$\frac{1}{n} \sum_{i=1}^n S(Q, X_i) \leq \frac{1}{n} \sum_{i=1}^n S(P, X_i)$$

Ha most a mintaelemszámmal a végtelenbe tartunk, akkor a nagy számok erős törvénye szerint,

$$\frac{1}{n} \sum_{i=1}^n S(P, X_i) \rightarrow S(P, Q) \quad \text{és} \quad \frac{1}{n} \sum_{i=1}^n S(Q, X_i) \rightarrow S(Q, Q)$$

így tehát épp a lényegesség definíciója miatt 1-hez tartó valószínűséggel nem fogunk tévedni és mindig a helyes modellt fogjuk kiválasztani. Éppen ezért definiáltuk így a lényeges szkórokat, mert ez a tulajdonság biztosítja, hogy a helyes modellt fogjuk kiválasztani, ha elegendő

adat áll a rendelkezésünkre. Na de nekünk az első mintából becsülni kellett az ismeretlen paraméter(ek)et. Ebben az esetben is át lehet menteni az előző eredményt? Legalább a logaritmikusság esetében? Ehhez tegyük fel, hogy a Q valódi és a P prediktív eloszlás is az exponenciális család tagja, azaz abszolút folytonosak és a sűrűségfüggvények, ahol f, g, ϕ, φ folytonos függvények:

$$p(x|\theta) = f(\theta)k(x) \exp\left(\sum_{j=1}^m \phi_j(\theta)t_j(x)\right)$$

$$q(x|\vartheta) = g(\vartheta)l(x) \exp\left(\sum_{j=1}^m \varphi_j(\vartheta)s_j(x)\right)$$

Becsüljük a paramétereket maximum likelihood: $\hat{p} = p(x|\hat{\theta})$. Ekkor

$$\frac{1}{n} \sum_{i=1}^n LS(\hat{p}, X_i) = \log f(\hat{\theta}) + \overline{\log k(X)} + \sum_{j=1}^m \phi_j(\hat{\theta}) \frac{1}{n} \sum_{i=1}^n t_j(x)$$

Ha most a mintaelemszámmal végtelenbe tartunk, akkor mivel f és ψ folytonosak ez a ugyanoda tart, mint az az átlagszkór, ahol ismertnek tekintjük a paramétereket:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n LS(\hat{p}, X_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n LS(p, X_i) = LS(P, Q)$$

Teljesen hasonlóan kapjuk, hogy a másik oldal esetében

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n LS(\hat{q}, X_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n LS(q, X_i) = LS(Q, Q)$$

Így tehát ekkor ebben az esetben is fennáll, hogy 1-hez tartó valószínűséggel nem fogunk tévedni.

Vizsgáljuk meg az előző példát Bayesi környezetben is: hasonlítsuk össze az Exponenciális modellt, ezzel a szintén λ várható értékű Gamma modellel a Bayes-faktor segítségével. Írjunk elő egy a priori eloszlást a λ ismeretlen paraméterre. Mindkét esetben a konjugált prior Gamma lesz, így a számítások megkönnyítése végett használjuk ezt.

Tehát legyen λ a priori eloszlása $\Gamma_{a,b}$, azaz $p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$

A két modell ekkor

$M_1 : Exp(\lambda)$ likelihood $\Gamma_{a,b}$ priorral

$M_2 : \Gamma_{\alpha,\alpha\lambda}$ likelihood $\Gamma_{a,b}$ priorral

A megfelelő sűrűségfüggvények pedig:

$$p_1(x|\lambda) = \lambda e^{-\lambda x}, \quad p_2(x|\lambda) = \frac{(\alpha\lambda)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha\lambda x}$$

Felírva a Bayes-faktort a két modellre:

$$B_1(x) = \frac{\int_0^\infty p_1(x|\lambda)p(\lambda)d\lambda}{\int_0^\infty p_2(x|\lambda)p(\lambda)d\lambda} = \frac{\int_0^\infty \lambda^n e^{-\lambda n\bar{x}} \lambda^{a-1} e^{-b\lambda} d\lambda}{\int_0^\infty \frac{(\alpha\lambda)^{\alpha n}}{\Gamma(\alpha)^n} \prod_{i=1}^n x_i^{\alpha-1} e^{-\alpha\lambda n\bar{x}} \lambda^{a-1} e^{-b\lambda} d\lambda} =$$

$$= \left(\frac{\Gamma(\alpha)}{\alpha^\alpha}\right)^n \frac{1}{\prod_{i=1}^n x_i^{\alpha-1}} \frac{\int_0^\infty \lambda^{a+n-1} e^{-\lambda(b+n\bar{x})} d\lambda}{\int_0^\infty \lambda^{a+\alpha n-1} e^{-\lambda(b+\alpha n\bar{x})} d\lambda} =$$

$$= \left(\frac{\Gamma(\alpha)}{\alpha^\alpha} \right)^n \frac{1}{\prod_{i=1}^n x_i^{\alpha-1}} \frac{\Gamma(a+n)}{\Gamma(a+\alpha n)} \frac{(\alpha n \bar{x} + b)^{a+\alpha n}}{(n \bar{x} + b)^{a+n}}$$

Ha most az a hiperparaméterrel végtelenhez tartunk, akkor az $\alpha < 1$ és az $\alpha > 1$ eseteket meg kell különböztetnünk. Előbbinél az a -tól függő

$$\frac{\Gamma(a+n)}{\Gamma(a+\alpha n)} \frac{(\alpha n \bar{x} + b)^{a+\alpha n}}{(n \bar{x} + b)^{a+n}}$$

kifejezés 0-hoz, míg a másik esetben szintén végtelenhez tart, persze rögzített n mintaelemszám mellett. Vagyis $\alpha < 1$ esetén, ahogy növeljük az a priori eloszlás szórásnégyzetét, úgy nő a Gamma modell evidenciája is az Exponenciálissal szemben. Az $\alpha > 1$ esetben épp fordítva, mindig meg tudjuk választani az a paraméter értékét úgy, hogy az Exponenciális modellt fogjuk elfogadni a Bayes faktor értéke alapján.¹³ Tehát ismételten megállapíthatjuk, hogy mennyire érzékeny a Bayes faktor az a priori hiperparaméterek megválasztására.

¹³Ez igaz marad, akkor is, ha az előbbi helyett a $\Gamma_{a,ac}$ a priori eloszlást használjuk.

Köszönetnyilvánítás

Köszönettel tartozom témavezetőmnek, Arató Miklósnak, az érdekes témajavaslatért, és a rendszeresen konzultációkon nyújtott segítségéért. Számptalan szakmai és formai javaslatával sokban segítette a dolgozat elkészültét.

Irodalomjegyzék

- [1] A. A. Borovkov - Matematikai Statisztika, *Typotex kiadó, Budapest, 1999*
- [2] Tilmann Gneiting and Adrian E. Raftery - Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, **102**, 359-378
- [3] Tilmann Gneiting, Fadoua Balabdaoui, Adrian E. Raftery - Probabilistic forecasts, calibration and sharpness, *Technical Report no. 483, Department of Statistics, University of Washington*
- [4] David Draper, Milovan Krnjajić, Thanasis Kottas - Bayesian Model Specification, *Department of Applied Mathematics and Statistics, University of California, Santa Cruz*
- [5] Claudia Czado, Tilmann Gneiting, Leonhard Held - Predictive Model Assessment for Count Data, *Biometrics*, to appear
- [6] Susanne Gschlößl, Claudia Czado - Spatial modelling of claim frequency and claim size in non-life insurance, *Scandinavian Actuarial Journal 2007*, **3**, 202-225
- [7] Kabaila P., He Z. - On prediction intervals for conditionally heteroscedastic processes, *Journal of Time Series Analysis*, **22**, 725-731
- [8] Cervera J. L., Muñoz J. - Proper scoring rules for fractiles, *Bayesian Statistics 5, Oxford University Press*
- [9] José M. Bernardo, Adrian F. M. Smith - Bayesian Theory, *John Wiley, New York*
- [10] B. Walsh - Introduction to Bayesian Analysis, *Lecture Notes for EEB 596z*
- [11] Roman Krzysztofowicz, Ashley A. Sigrest - Calibration of Probabilistic Quantitative Precipitation Forecasts, *Weather and Forecasting*, **14**, 427-442
- [12] Székely J. Gábor - ϵ -Statistics: The energy of statistical samples. *Technical Report no. 2003-16, Department of Mathematics and Statistics, Bowling Green State University, Ohio*
- [13] Good I. J. - Rational decisions, *Journal of the Royal Statistical Society*, **14**, 107-114
- [14] Winkler R. L. - Evaluating probabilities: Asymmetric scoring rules. *Management Science*, **40**, 1395-1405