Maximum likelihood

Rövid bevezetés

Medvegyev Péter

Corvinus

2011

Maximum és qvázi maximum likelihood

A maximum likelihood becslés, MLE, matematikai tárgyalása során fel szokás tenni, hogy a becslés csak egy véges paramétertartományból veheti az értékét és a helyes paraméter, és így a helyes eloszlás része a paraméter halmaznak. Ha a valós eloszlás nem található meg a paraméterek között, vagy a minta elemek nem függetlenek, akkor azt mondjuk, hogy a modell rosszul lett specifikálva. Ilyenkor nem maxium likelihood, hanem kvázi maximum likelihood becslésről beszélünk. (QMLE estimator.) Mit lehet ilyenkor mondani? Ehhez először a hagyományos MLE becslés tulajdonságait tárgyaljuk:

Jensen-egyenlőtlenség

Theorem

Ha $f : \mathbb{R} \to \mathbb{R}$ *konvex, akkor tetszőleges* ξ *változóra*

$$f\left(\mathbf{E}\left(\xi\right)\right) \leq \mathbf{E}\left(f\left(\xi\right)\right)$$

feltéve, hogy az egyenlőtlenségben szereplő véges vagy végtelen várható értékek léteznek.

Megjegyezzük, hogy konvex függvényekre mindig létezik a $\pm\infty$-ben vett, esetlegesen végtelen értékű, határérték, így ha az $\mathbf{E}\left(\xi\right)$ létezik, akkor a bal oldali kifejezés értelmes.

Bizonyítás

1. Tegyük fel először, hogy az $\mathbf{E}(\xi)$ és az $f(\mathbf{E}(\xi))$ véges. Tekintsünk egy az

$$(\mathbf{E}(\xi), f(\mathbf{E}(\xi)))$$

ponton átmenő olyan $ax + b$ alakú egyenest, hogy minden x számra

$$f(x) \geq a \cdot x + b.$$

Az f függvény konvexitása miatt ilyen egyenes biztosan van. Ha f deriválható, akkor a deriváltja éppen ilyen, ha nem akkor a konvexitás miatt létező f'_+ jobb, illetve f'_- bal oldali deriváltak, illetve a köztük haladó bármely másik egyenes megfelelő.

Bizonyítás

De ekkor az alapul vett Ω valószínűségi mező minden $\omega \in \Omega$ kimenetelére

$$f\left(\xi\left(\omega\right)\right) \geq a \cdot \xi\left(\omega\right) + b.$$

Mindkét oldalt integrálva, az $ax + b$ egyenes választása alapján

$$\mathbf{E}\left(f\left(\xi\right)\right) \geq a \cdot \mathbf{E}\left(\xi\right) + b \cdot \mathbf{E}\left(1\right) = f\left(\mathbf{E}\left(\xi\right)\right).$$

Bizonyítás

2. Ha $\mathbf{E}(f(\xi)) = \infty$, az állítás evidens.

Bizonyítás

3. Tegyük fel, hogy $\mathbf{E}(\xi) = \infty$. Ha létezik t, hogy $f'_-(t) > 0$, akkor a konvexitásból következő

$$f(\xi) \geq f(t) + f'_-(t) \cdot (\xi - t)$$

egyenlőtlenséget integrálva és kihasználva, hogy az integrálok léteznek

$$\mathbf{E}(f(\xi)) \geq f(t) + f'_-(t) \cdot (\mathbf{E}(\xi) - t) = \infty,$$

tehát az egyenlőtlenség teljesül. Ha minden t-re $f'_-(t) \leq 0$, akkor az f monoton csökken, vagyis $f(\infty) \leq f(\xi)$ és ezért

$$\mathbf{E}(f(\xi)) \geq f(\infty) = f(\mathbf{E}(\xi)).$$

Bizonyítás

4. Az $\mathbf{E}(\xi) = -\infty$, analóg, csak az $f'_{-}(t) < 0$ esetből kell kiindulni.

Bizonyítás

5. Végezetül ha $\mathbf{E}(f(\xi)) = -\infty$, és van olyan t, amelyre $f'_-(t) > 0$, akkor az imént említett módon $\mathbf{E}(\xi) = -\infty$, és az egyenlőtlenség a már bemutatott módon indokolható. Ha $f'_- \leq 0$, és van olyan t, hogy $f'_-(t) < 0$, akkor egyrészt az f csökken, másrészt a már korábban elmondottak miatt $\mathbf{E}(\xi) = \infty$, vagyis $f(\mathbf{E}(\xi)) \leq f(\xi)$. Ha $f'_- = 0$, akkor az f véges konstans, és az $\mathbf{E}(f(\xi)) = -\infty$ nem teljesülhet.

Megjegyzés

Az egyenlőtlenségben explicite feltettük, hogy az $\mathbf{E}(\xi)$ és az $\mathbf{E}(f(\xi))$ várható értékek léteznek.

Theorem

Ha $f : \mathbb{R} \to \mathbb{R}$ *konvex és az* $\mathbf{E}(\xi)$ *várható érték véges, akkor*

$$\mathbf{E}\left(f^{-}(\xi)\right) < \infty,$$

következésképpen az $\mathbf{E}(f(\xi))$ *véges vagy végtelen várható érték létezik.*

Bizonyítás: A konvexitás miatt ismételten

$$f(t) \geq f(0) + f'_{-}(0) \cdot t,$$

$$-f(t) \leq -f(0) - f'_{-}(0) \cdot t,$$

$$f^{-}(t) \circeq \max(-f(t), 0) \leq |f(0)| + \left|f_{-}^{-}(0)\right| \cdot |t|,$$

$$\mathbf{E}\left(f^{-}(\xi)\right) \leq |f(0)| + \left|f_{-}^{-}(0)\right| \cdot \mathbf{E}(|\xi|) < \infty.$$

Az elv indoklása

Legyen μ tetszőleges mérték és legyenek p és q a μ mérték szerinti sűrűségfüggvények. A μ lehet például a számláló mérték. Ilyenkor a diszkrét eloszlásokat kezeljük és az integrálok egyszerű összegek. A logaritmus függvény konkáv, így a Jensen-egyenlőtlenség szerint

$$\begin{aligned}\int_\Omega p\cdot(\ln q-\ln p)\,d\mu &= \int_\Omega p\cdot\ln\frac{q}{p}d\mu = \mathbf{E}\left(\ln\frac{q}{p}\right)\leq \ln\left(\mathbf{E}\left(\frac{q}{p}\right)\right)= \\ &= \ln\left(\int_\Omega p\cdot\frac{q}{p}d\mu\right)=\ln\left(\int_\Omega q d\mu\right)=\ln 1=0.\end{aligned}$$

Ebből következően

$$\int_\Omega p\cdot\ln q d\mu \leq \int_\Omega p\cdot\ln p d\mu, \tag{1}$$

feltéve, hogy mind a két oldal véges.

Az elv indoklása

Az egyenlőtlenség alkalmazásaként tegyük fel, hogy a $\xi_1, \xi_2, \ldots, \xi_n$ változók azonos eloszlásúak, és a közös sűrűségfüggvényük legyen $f(x, \theta)$, ahol θ egy ismeretlen paraméter amely értékét a $(\xi_k)_{k=1}^n$ változók megfigyelése alapján próbáljuk megbecsülni. Képezzük az

$$L(\xi_1, \ldots, \xi_n, \theta) \stackrel{\circ}{=} \prod_{k=1}^{n} f(\xi_k, \theta)$$

likelihood függvényt. Ha a ξ_k változók függetlenek, akkor a nagy számok törvénye alapján

$$\frac{\ln L(\xi_1, \ldots, \xi_n, \theta)}{n} = \frac{\sum_{k=1}^n \ln f(\xi_k, \theta)}{n} \to \mathbf{E}(\ln f(\xi, \theta)),$$

ahol ξ tetszőleges olyan változó, amely eloszlása megegyezik a ξ_k változók közös eloszlásával.

elv indoklása

Ha θ_0 a θ paraméter valódi értéke, akkor a transzformált valószínűségi változók várható értékére vonatkozó formula alapján

$$\mathbf{E}\left(\ln f\left(\xi,\theta\right)\right) \circeq \int_{\Omega} \ln f\left(\xi,\theta\right) d\mathbf{P} = \int_{\Omega} \ln f\left(x,\theta\right) f\left(x,\theta_0\right) d\mu\left(x\right).$$

Az (1) alapján

$$\begin{aligned} \mathbf{E}\left(\ln f\left(\xi,\theta\right)\right) &= \int_{\Omega} \ln f\left(x,\theta\right) f\left(x,\theta_0\right) d\mu\left(x\right) \leq \\ &\leq \int_{\Omega} \ln f\left(x,\theta_0\right) f\left(x,\theta_0\right) d\mu\left(x\right) \leq \mathbf{E}\left(\ln f\left(\xi,\theta_0\right)\right). \end{aligned}$$

Az elv indoklása

Ebből következően a θ_0 ismeretlen paraméter $\widehat{\theta}$ becslését úgy célszerű meghatározni, hogy maximalizáljuk a várható értéket közelítő

$$\frac{\ln L\left(\xi_1, \ldots, \xi_n, \theta\right)}{n}$$

átlag értékét a θ paraméter szerint. Ez ekvivalens avval, hogy adott n esetén a θ paraméter szerint maximalizáljuk az L likelihood függvényt.

Becslés hibája

Mekkora a becslés hibája, vagyis mekkora a $\widehat{\theta} - \theta_0$ eltérés? Tekintsük az $\ln L$ függvény Taylor-kifejtését a θ_0 helyes érték körül:

$$\begin{aligned}\ln L(\theta) &= \ln L(\theta_0) + \frac{d \ln L}{d\theta}(\theta_0)(\theta - \theta_0) + \frac{1}{2}\frac{d^2 \ln L}{d\theta^2}(\theta_0)(\theta - \theta_0)^2 + \\ &\quad + R_3 (\theta - \theta_0)^3 .\end{aligned}$$

A kifejtést mint polinomot θ szerint deriválva, felhasználva, hogy az első tag kontans

$$\frac{d}{d\theta}\ln L(\theta) = \frac{d \ln L}{d\theta}(\theta_0) + \frac{d^2 \ln L}{d\theta^2}(\theta_0)(\theta - \theta_0) + 3R_3(\theta - \theta_0)^2 .$$

Becslés hibája

A θ helyébe a $\widehat{\theta}$ becslést behelyettesítve és a T_3 új jelölést bevezetve

$$\frac{d}{d\theta}\ln L\left(\widehat{\theta}\right)=\frac{d\ln L}{d\theta}\left(\theta_0\right)+\frac{d^2\ln L}{d\theta^2}\left(\theta_0\right)\left(\widehat{\theta}-\theta_0\right)+T_3\left(\widehat{\theta}-\theta_0\right)^2.$$

A $\widehat{\theta}$ értékét a maximálással becsültük, a maximumban a derivált nulla, így a jobb oldalon $\frac{d}{d\theta}\ln L\left(\widehat{\theta}\right)=0$, vagyis

$$0=\frac{d\ln L}{d\theta}\left(\theta_0\right)+\frac{d^2\ln L}{d\theta^2}\left(\theta_0\right)\left(\widehat{\theta}-\theta_0\right)+T_3\left(\widehat{\theta}-\theta_0\right)^2.$$

Ebből elemi átrendezéssel

$$\left(\widehat{\theta}-\theta_0\right)=-\left(\frac{d\ln L}{d\theta}\left(\theta_0\right)+T_3\left(\widehat{\theta}-\theta_0\right)^2\right)\left(\frac{d^2\ln L}{d\theta^2}\left(\theta_0\right)\right)^{-1}.$$

Becslés hibája

A két oldalt $\sqrt{n}$-bel beszorozva

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) = -\frac{n}{\sqrt{n}}\left(\frac{d\ln L}{d\theta}(\theta_0) + T_3\left(\widehat{\theta}-\theta_0\right)^2\right)\left(\frac{d^2\ln L}{d\theta^2}(\theta_0)\right)^{-1} =$$

$$= -\left(\frac{\frac{d\ln L}{d\theta}(\theta_0) + T_3\left(\widehat{\theta}-\theta_0\right)^2}{\sqrt{n}}\right)\left(\frac{\frac{d^2\ln L}{d\theta^2}(\theta_0)}{n}\right)^{-1}.$$

Első tag

$$\frac{\frac{d\ln L}{d\theta}\left(\theta_0\right)}{\sqrt{n}} = \frac{\sum_{k=1}^{n}\frac{d}{d\theta}\ln f\left(\xi_k,\theta_0\right)}{\sqrt{n}}.$$

A centrális határeloszlás tételét akarjuk használni, de ehhez szükségünk van az összeadandók várható értékre és szórására. Megmutatjuk, hogy az

$$\eta_k \circeq \frac{d}{d\theta}\ln f\left(\xi_k,\theta_0\right)$$

változók közös várható értéke nulla.

Első tag, várható érték

$$
\begin{aligned}
\mathbf{E}\left(\eta_k\right) &\stackrel{\circ}{=} \mathbf{E}\left(\frac{d}{d\theta}\ln f\left(\xi_k,\theta_0\right)\right) = \\
&= \mathbf{E}\left(\frac{1}{f\left(\xi_k,\theta_0\right)}\frac{d}{d\theta}f\left(\xi_k,\theta_0\right)\right) = \\
&= \int_{\mathbb{R}}\frac{1}{f\left(x,\theta_0\right)}\frac{d}{d\theta}f\left(x,\theta_0\right)f\left(x,\theta_0\right)d\mu = \\
&= \int_{\mathbb{R}}\frac{d}{d\theta}f\left(x,\theta_0\right)d\mu = \\
&= \frac{d}{d\theta}\int_{\mathbb{R}}f\left(x,\theta\right)d\mu = \frac{d}{d\theta}1 = 0.
\end{aligned}
$$

Első tag, regularitási tulajdonság

Vegyük észre, hogy a számolás során kihasználtuk, hogy θ_0 a paraméter helyes értéke, így az **E** várható érték az $f(x, \theta_0)$ szerinti integrálással számolható. Ugyancsak felhasználtuk, hogy a deriválás és az integrálás felcserélhető, amely nyilván általában nem teljesül. Ez megszorítást jelent a számba jöhető $f(x, \theta)$ párokra és nyilván a konkrét alkalmazásokban ellenőrízni kell ezt a regularitási tulajdonságot.

Első tag, szórás

Számoljuk ki az η_k változók közös szórását. Mivel a várható érték nulla elég kiszámolni a második momentumot. Egyrészt a már felhasznált regularitási szabály alapján

$$\frac{d}{d\theta}\int_{\mathbb{R}}\frac{d}{d\theta}\ln f(x,\theta)\, f(x,\theta)\, d\mu =$$

$$= \frac{d}{d\theta}\int_{\mathbb{R}}\left(\frac{1}{f(x,\theta)}\frac{d}{d\theta}f(x,\theta)\right) f(x,\theta)\, d\mu =$$

$$= \frac{d}{d\theta}\int_{\mathbb{R}}\frac{d}{d\theta}f(x,\theta)\frac{1}{f(x,\theta)}f(x,\theta)\, d\mu = \frac{d}{d\theta}\int_{\mathbb{R}}\frac{d}{d\theta}f(x,\theta)\, d\mu =$$

$$= \frac{d^2}{d\theta^2}\int_{\mathbb{R}} f(x,\theta)\, d\mu = \frac{d^2}{d\theta^2}1 = 0.$$

Első tag, szórás

Az integrálba bederiválva a szorzat deriválási szabálya szerint ugyanez

$$0 = \frac{d}{d\theta} \int_{\mathbb{R}} \frac{d}{d\theta} \ln f(x,\theta) f(x,\theta)\, d\mu =$$
$$= \int_{\mathbb{R}} \frac{d}{d\theta} \left(\frac{d}{d\theta} \ln f(x,\theta) f(x,\theta) \right) d\mu =$$
$$= \int_{\mathbb{R}} \frac{d^2}{d\theta^2} \ln f(x,\theta) f(x,\theta)\, d\mu + \int_{\mathbb{R}} \frac{d}{d\theta} \ln f(x,\theta) \frac{d}{d\theta} f(x,\theta)\, d\mu.$$

Első tag, szórás

A már többször használt

$$\frac{d}{d\theta} f(x,\theta) = \frac{d}{d\theta} \ln f(x,\theta) f(x,\theta)$$

azonosság szerint, ugyanis az előző slide-on az összeg nulla.

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{d^2}{d\theta^2} \ln f(x,\theta) f(x,\theta)\, d\mu = \\
&= -\int_{\mathbb{R}} \frac{d}{d\theta} \ln f(x,\theta) \frac{d}{d\theta} f(x,\theta)\, d\mu = \\
&= -\int_{\mathbb{R}} \frac{d}{d\theta} \ln f(x,\theta) \frac{d}{d\theta} \ln f(x,\theta) f(x,\theta)\, d\mu = \\
&= -\int_{\mathbb{R}} \left(\frac{d}{d\theta} \ln f(x,\theta) \right)^2 f(x,\theta)\, d\mu .
\end{aligned}$$

Első tag, szórás

Ha most $\theta \overset{\circ}{=} \theta_0$, akkor az $f(x, \theta_0)$ a valódi sűrűségfüggvény, így az integrálok a transzformált változók várható értékét adják:

$$\mathbf{E}\left(\left(\frac{d}{d\theta}\ln f(x, \theta_0)\right)^2\right) = -\mathbf{E}\left(\frac{d^2}{d\theta^2}\ln f(x, \theta_0)\right),$$

vagyis

$$\mathbf{E}\left(\eta^2\right) = \mathbf{D}^2(\eta) = \mathbf{D}^2(\eta_k) \overset{\circ}{=} \mathbf{D}^2\left(\frac{d}{d\theta}\ln f(\xi_k, \theta_0)\right) =$$

$$= \mathbf{E}\left(\left(\frac{d}{d\theta}\ln f(x, \theta_0)\right)^2\right) =$$

$$= -\mathbf{E}\left(\frac{d^2}{d\theta^2}\ln f(\xi_k, \theta_0)\right) \overset{\circ}{=} \varphi.$$

Első tag, összefoglalás

Ebből a centrális határeloszlás tétel szerint

$$\begin{aligned}
\lim_{n\to\infty}\frac{\frac{d\ln L}{d\theta}\left(\theta_0\right)}{\sqrt{n}} &= \lim_{n\to\infty}\frac{\sum_{k=1}^{n}\frac{d}{d\theta}\ln f\left(\xi_k,\theta_0\right)}{\sqrt{n}} = \\
&= \lim_{n\to\infty}\frac{\sum_{k=1}^{n}\left(\eta_k-\mathbf{E}\left(\eta_k\right)\right)}{\sqrt{n}} = N\left(0,\mathbf{D}\left(\eta\right)\right) = \\
&= N\left(0,\sqrt{\varphi}\right).
\end{aligned}$$

Második tag

Ugyanakkor a nagy számok törvénye és a már belátott

$$\int_{\mathbb{R}} \frac{d^2}{d\theta^2} \ln f(x, \theta_0) f(x, \theta_0) \, d\mu = \mathbf{E}\left(\frac{d^2}{d\theta^2} \ln f(\xi, \theta_0)\right) = -\varphi$$

szerint

$$\begin{aligned} \lim_{n\to\infty} \frac{\frac{d^2 \ln L}{d\theta^2}(\theta_0)}{n} &= \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \frac{d^2 \ln f(\xi_k, \theta)}{d\theta^2}(\theta_0)}{n} = \\ &= \mathbf{E}\left(\frac{d^2}{d\theta^2} \ln f(\xi, \theta_0)\right) = -\varphi. \end{aligned}$$

Harmadik tag

Tegyük fel, hogy a becslés "elég aszimptotikusan konzisztens", vagyis a Taylor-közelítés maradékjára

$$\lim_{n\to\infty} \frac{T_3\left(\widehat{\theta}_n - \theta_0\right)^2}{\sqrt{n}} = 0.$$

Ez teljesül például akkor, ha a harmadik derivált a becslés környezetében korlátos és a paramétertér, vagyis a θ lehetséges értékei halmaza például szintén korlátos. De számos más egyéb a konkrét helyzetben ellenőrizendő feltétel esetén teljesül.

Összefoglalás

Ilyenkor eloszlásban

$$\lim_{n\to\infty} \sqrt{n}\left(\widehat{\theta}-\theta_0\right) =$$

$$= \lim_{n\to\infty} = -\left(\frac{\frac{d \ln L}{d\theta}\left(\theta_0\right) + T_3\left(\widehat{\theta}-\theta_0\right)^2}{\sqrt{n}}\right)\left(\frac{\frac{d^2 \ln L}{d\theta^2}\left(\theta_0\right)}{n}\right)^{-1} =$$

$$= -N\left(0, \sqrt{\varphi}\right)\left(-\varphi^{-1}\right) = N\left(0, \sqrt{\varphi^{-1}}\right).$$

A szórás becslése

Emlékeztetünk, hogy

$$\varphi \mathrel{\hat{=}} -\mathbf{E}\left(\frac{d^2}{d\theta^2}\ln f\left(\xi_k,\theta_0\right)\right)=\mathbf{E}\left(\left(\frac{d}{d\theta}\ln f\left(x,\theta_0\right)\right)^2\right).$$

A nagy számok törvénye szerint

$$\begin{aligned}\mathbf{E}\left(\frac{d^2}{d\theta^2}\ln f\left(\xi_k,\theta_0\right)\right) &\approx \frac{\sum_{k=1}^{n}\frac{d^2}{d\theta^2}\ln f\left(\xi_k,\theta_0\right)}{n}\approx \\ &\approx \frac{\sum_{k=1}^{n}\frac{d^2}{d\theta^2}\ln f\left(\xi_k,\widehat{\theta}\right)}{n}.\end{aligned}$$

Nem jó a modell

Mit tudunk mondani akkor, ha az alapul vett modell ténylegesen nem azonos a valóságos modellel. Természetesen az alkalmazásokban ez szinte mindig teljesül, tehát ez tekinthető a standard esetnek. Legyen G a tényleges eloszlás és jelölje g a sűrűségfüggvényét. A maximum likelihood becslés az a $\widehat{\theta}$, amely minimalizálja a Kullback–Leibler távolságot:

$$\begin{aligned} I(g, f(\cdot, \theta)) &= \mathbf{E}\left(\ln \frac{g(\xi)}{f(\xi, \theta)}\right) = \\ &= \int_{\mathbb{R}} \ln g(x)\, dG(x) - \int_{\mathbb{R}} \ln f(x, \theta)\, dG(x). \end{aligned}$$

Nem jó a modell

A távolság tartalmának megértése céljából vegyük észre, hogy a logaritmus alatt álló kifejezés, minden rögzített θ esetén éppen a $\frac{d\mathbf{P}}{d\mathbf{Q}}$ Radon–Nikodym derivált, ahol $\mathbf{P}$ a helyes eloszlás, a $\mathbf{Q}$ pedig az f-hez tartozó eloszlás. Ennek megfelelően a távolság éppen $\mathbf{E}\left(\ln \frac{d\mathbf{P}}{d\mathbf{Q}}\right)$. Érdemes hangsúlyozni, hogy az induló feltételek szerint a sűrűségfüggvények léteznek. és ugyancsak implicite feltettük,hogy a g/f hányados értelmes, vagyis nem fordul elő az, hogy a g valamilyen pontban pozitív és ott az f nulla. Ez azt jelenti, hogy a helyes eloszlás abszolút folytonos a becslésben használtakra nézve. Vagyis nem fordulhat elő, hogy egy esemény értéke nulla a becsléshez használt családban és mégis a tényleges valószínűsége nem nulla. Az utóbbi azt jelentené, hogy releváns információt a becslés során nem veszünk figyelembe.

A távolság értelmezése

Általános esetben

$$I\left(\mathbf{P} \mid \mathbf{Q}\right) \mathrel{\hat{=}} \mathbf{E}\left(\ln \frac{d\mathbf{P}}{d\mathbf{Q}}\right).$$

Vegyük még észre, hogy az $A = \left\{\frac{d\mathbf{P}}{d\mathbf{Q}} = 0\right\}$ esemény valószínűsége nulla, ugyanis

$$\mathbf{P}\left(A\right) = \int_{\Omega} \chi_A d\mathbf{P} = \int_{\Omega} \chi_A \frac{d\mathbf{P}}{d\mathbf{Q}} d\mathbf{Q} = \int_{\Omega} 0 d\mathbf{Q} = 0,$$

így a várható érték alatti logaritmus majdnem mindenhol értelmes, így az $\mathbf{E}\left(\ln \frac{d\mathbf{P}}{d\mathbf{Q}}\right)$ minden esetben értelmes, feltéve ha $\mathbf{P} \ll \mathbf{Q}$. A logaritmus függvény indoklása a következő: Ha a $\mathbf{P}$ két független valószínűség változó megfigyeléséből ered, vagyis $\mathbf{P} = \mathbf{P}_1\mathbf{P}_2$, akkor :az együttes eloszlás által adott információ éppen az információk összege, vagyis ilyenkor a szorzatmértékek információtartalma legyen az információk összege.

A becslés konzisztens, de nem a jó paramétert becsli

A QMLE estimator akárcsak a klasszikus esetben konzisztens becslése annak a θ_*-nak, amely minimalizáltja a KL-távolságot. Ezonban ez az érték nem feltétlenül konzisztens becslése a keresett paraméternek. Például, ha a szórást akarjuk becsülni, de a „közelítő" eloszlásokban a várható érték, helytelenül, nullára van rögzítve, akkor a becslés a minta elemszámának növelésével sem közelíthet a tényleges szórás értékhez. Ugyanakkor előfordulhat az is, hogy a közelítő eloszlásra tett helytelen hipotézis ellenére a becslés konzisztens. Tipikus példa a normalitás feltétele a várható érték becslésére.

A hiba becslése

Az előző levezetésben a becslési hiba aszimptotikus normalitásának levezetésekor beláttuk, hogy

$$\mathbf{E}\left(\left(\frac{d}{d\theta}\ln f\left(x,\theta_0\right)\right)^2\right)=-\mathbf{E}\left(\frac{d^2}{d\theta^2}\ln f\left(\xi,\theta_0\right)\right)\doteq\varphi.$$

Ha a θ paramétertér több dimenziós, akkor ez

$$\mathbf{E}\left(\frac{\partial}{\partial\theta_i}\ln f\left(\boldsymbol{\xi},\theta_0\right)\frac{\partial}{\partial\theta_j}\ln f\left(\boldsymbol{\xi},\theta_0\right)\right)=-\mathbf{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f\left(\boldsymbol{\xi},\theta_0\right)\right)\doteq\Phi$$

alakban általánosítható. Vagyis a Hesse-mátrix várható értékének mínusz egyszerese éppen az első deriváltak diadikus szorzatának várható értéke. Emlékeztetünk, hogy maximumban a Hesse-matrix negatív semidefinite, vagyis a Φ pozitív semidefinite.

A hiba becslése

Vezessük be az

$$A(\theta) \stackrel{\circ}{=} \mathbf{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f(\boldsymbol{\xi},\theta)\right), \quad B(\theta) \stackrel{\circ}{=} \mathbf{E}\left(\frac{\partial}{\partial\theta_i}\ln f(\boldsymbol{\xi},\theta)\frac{\partial}{\partial\theta_j}\ln f(\boldsymbol{\xi},\theta)\right)$$

mátrixokat. Legyen

$$C(\theta) = A^{-1}(\theta)\, B(\theta)\, A^{-1}(\theta).$$

.A $B(\theta)$ pozitív semidefinite, az $A(\theta)$ így az $A^{-1}(\theta)$ szimmetrikus, így a $C(\theta)$ is pozitív semidefinite. Megmutatható, hogy a becslési hiba aszimptotikus normalitására vonatkozó szabály a következő

$$\sqrt{n}\,(\theta_n - \theta_*) \xrightarrow{w} N(0, C(\theta_*)),$$

ahol a θ_* a legkisebb Kullback–Leibler távolságot biztosító paraméter érték együttes.

A hiba becslése

Ha $A(\theta) = -B(\theta)$, vagyis $A(\theta) + B(\theta) = 0$, akkor éppen

$$\begin{aligned} C(\theta) &\overset{\circ}{=} A^{-1}(\theta) B(\theta) A^{-1}(\theta) = \\ &= A^{-1}(\theta)(-A(\theta)) A^{-1}(\theta) = -A^{-1}(\theta) \overset{\circ}{=} \Phi. \end{aligned}$$

A két érték azonban nem azonos.

A normalitás esetén való példa

Példaként tekintsük a normális eloszlás paramétereinek becslését. Ekkor

$$
\begin{aligned}
l(x,\mu,\sigma) &= \ln\varphi_{\mu,\sigma}(x) = -\ln\sqrt{2\pi} - \ln\sigma - \frac{(x-\mu)^2}{2\sigma^2} = \\
&= \ln\varphi_{\mu,\sigma}(x) = -\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2} \\
\ln L &= -n\ln\sqrt{2\pi} - \frac{1}{2}n\ln\sigma^2 - \sum_{k=1}^{n}\frac{(\xi_k-\mu)^2}{2\sigma^2}..
\end{aligned}
$$

A várható érték és szórás becslése

Az optimalitást megadó statcionaritási egyenletek n elemű minta esetén μ és $V \doteq \sigma^2$ szerint véve

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \mu} = -2\sum_{k=1}^{n} \frac{\xi_k - \mu}{2\sigma^2} = n\frac{\overline{\xi}_n - \mu}{\sigma^2} \\ 0 &= \frac{\partial L}{\partial \sigma^2} = -\frac{1}{2}\frac{n}{\sigma^2} + \sum_k \frac{\left(\xi_k - \mu\right)^2}{2\sigma^4} = \\ &= -\frac{n}{2\sigma^2} + \frac{\sum_k \left(\xi_k - \overline{\xi}_n\right)^2 + n\left(\overline{\xi}_n - \mu\right)^2}{2\sigma^4} \end{aligned}$$

ahol

$$\overline{\xi}_n \doteq \frac{\sum_{k=1}^{n} \xi_k}{n}.$$

A várható érték és szórás becslése

Ugyanis a számláló

$$\sum_k \left(\xi_k - \overline{\xi}_n\right)^2 + n\left(\overline{\xi}_n - \mu\right)^2 =$$

$$= \sum_k \xi_k^2 - 2\sum_k \xi_k \overline{\xi}_n + n\overline{\xi}_n^2 + n\left(\overline{\xi}_n - \mu\right)^2 =$$

$$= \sum_k \xi_k^2 - 2\sum_k \xi_k \overline{\xi}_n + n\overline{\xi}_n^2 + n\overline{\xi}_n^2 + n\mu^2 - 2n\overline{\xi}_n\mu =$$

$$= \sum_k \xi_k^2 - 2n\overline{\xi}_n^2 + n\overline{\xi}_n^2 + n\overline{\xi}_n^2 + n\mu^2 - 2n\overline{\xi}_n\mu =$$

$$= \sum_k \xi_k^2 + n\mu^2 - 2n\overline{\xi}_n\mu = \sum_k \left(\xi_k - \mu\right)^2 .$$

A várható érték és szórás becslése

Az első egyenletből

$$\widehat{\mu}_n = \frac{\sum_{k=1}^{n} \xi_k}{n} \to \mathbf{E}(\xi) = \mu.$$

A μ becslését a másodikba betéve

$$0 = -\frac{n}{\sigma^2} + \frac{\sum_k \left(\xi_k - \overline{\xi}_n\right)^2}{\sigma^4}.$$

A σ-val beszorozva és átrendezve

$$\widehat{\sigma}_n^2 = \frac{\sum_k \left(\xi_k - \overline{\xi}_n\right)^2}{n} \to \mathbf{D}^2(\xi) = \sigma^2.$$

A két mátrix kiszámolása

Ugyanakkor az $l \circeq \ln \varphi_{\mu,\sigma}(x)$ második deriváltjait a μ és a $V \circeq \sigma^2$ paraméterek szerint véve

$$A\left(\mu,\sigma^2\right) \circeq \mathbf{E}\left(\begin{pmatrix} \partial^2 l/\partial\mu^2 & \partial^2 l/\partial\mu\partial V \\ \partial^2 l/\partial\mu\partial V & \partial^2 l/\partial V^2 \end{pmatrix}\right) =$$

$$= \mathbf{E}\begin{pmatrix} -1/\sigma^2 & -\left(\xi-\mu\right)/\sigma^4 \\ -\left(\xi-\mu\right)/\sigma^4 & 1/\left(2\sigma^4\right) - 2\frac{\left(\xi_k-\mu\right)^2}{2\sigma^6} \end{pmatrix} =$$

$$= \begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & 1/\left(2\sigma^4\right) - \frac{\sigma^2}{\sigma^6} \end{pmatrix} = \begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & -1/2\sigma^4 \end{pmatrix} = .$$

A két mátrix kiszámolása

Az elsőrendű parciális deriváltak

$$\begin{pmatrix} \partial l / \partial \mu & \partial l / \partial \sigma^2 \end{pmatrix} = \begin{pmatrix} \frac{x-\mu}{\sigma^2} & -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{pmatrix}$$

A diadilus szorzat várható értéke

$$B\left(\mu, \sigma^2\right) \stackrel{\circ}{=} \mathbf{E} \begin{pmatrix} \frac{(\xi-\mu)^2}{\sigma^4} & -\frac{\xi-\mu}{\sigma^4} + \frac{(\xi-\mu)^3}{2\sigma^6} \\ -\frac{\xi-\mu}{\sigma^4} + \frac{(\xi-\mu)^3}{2\sigma^6} & \left(\frac{1}{2\sigma^2} - \frac{(\xi-\mu)^2}{2\sigma^4}\right)^2 \end{pmatrix} =$$

$$= \mathbf{E} \begin{pmatrix} 1/\sigma^2 & \frac{(\xi-\mu)^3}{2\sigma^6} \\ \frac{(\xi-\mu)^3}{2\sigma^6} & \left(\frac{1}{2\sigma^2} - \frac{(\xi-\mu)^2}{2\sigma^4}\right)^2 \end{pmatrix}.$$

Ferdeség és kurtózis

Ha

$$\beta \circeq \mathbf{E}\left(\frac{\xi-\mu}{\mathbf{D}\left(\xi\right)}\right)^3$$

a ferdeségi mutató és

$$\kappa \circeq \mathbf{E}\left(\frac{\xi-\mu}{\mathbf{D}\left(\xi\right)}\right)^4$$

a kurtózis akkor

$$B\left(\mu,\sigma^2\right)=\begin{pmatrix} 1/\sigma^2 & \beta/2\sigma^3 \\ \beta/2\sigma^3 & \left(\kappa-1\right)/4\sigma^4 \end{pmatrix},$$

A két mátrix kiszámolása

ugyanis

$$\mathbf{E}\left(\left(-\frac{1}{2\sigma^2}+\frac{(\xi-\mu)^2}{2\sigma^4}\right)^2\right)=\frac{1}{4\sigma^4}\mathbf{E}\left(\left(-1+\left(\frac{\xi-\mu}{\sigma}\right)^2\right)^2\right)=$$

$$=\frac{1}{4\sigma^4}\left(1-2\mathbf{E}\left(\frac{\xi-\mu}{\sigma}\right)^2+\mathbf{E}\left(\frac{\xi-\mu}{\sigma}\right)^4\right)\stackrel{\circ}{=}$$

$$\stackrel{\circ}{=}\frac{1}{4\sigma^4}(1-2+\kappa)=\frac{1}{4\sigma^4}(\kappa-1).$$

A félrespecifikálás mértéke

Természetesen ha az eloszlás normális, akkor $\beta = 0$ és $\kappa = 3$. Ilyenkor $A = -B$. De a helytelen specifikáció miatt az egyenlőség nem teljesül. Számoljuk ki a $C\left(\mu, \sigma^2\right)$ mátrixot.

$$A^{-1}\left(\mu, \sigma^2\right) = \begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & -1/2\sigma^4 \end{pmatrix}^{-1} = \begin{pmatrix} -\sigma^2 & 0 \\ 0 & -2\sigma^4 \end{pmatrix}.$$

Így

$$C\left(\mu, \sigma^2\right) = A^{-1} B A^{-1} = \mathbf{E} \begin{pmatrix} \sigma^2 & \beta\sigma^3 \\ \beta\sigma^3 & (\kappa - 1)\,\sigma^4 \end{pmatrix}.$$

A félrespecifikálás mértéke

Az $A(\theta) \neq -B(\theta)$ reláció a modell félrespecifikálásának indikátora. Ez másképpen az $A(\theta) + B(\theta) = 0$ a modell helyes specifikációjának a jele. Ez normális eloszlás esetén a

$$\begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & -1/2\sigma^4 \end{pmatrix} + \begin{pmatrix} 1/\sigma^2 & \beta/2\sigma^3 \\ \beta/2\sigma^3 & (\kappa-1)/4\sigma^4 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & \beta/2\sigma^3 \\ \beta/2\sigma^3 & (\kappa-3)/4\sigma^4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

egyenlőségre redukálódik. Ez pedig ekvivalens a $\beta = 0$, $\kappa = 3$ egyenlőséggel, amely a normalitás közismert azon két tulajdonságával azonos, miszerint a ferdeségnek nullának, a kurtózisnak meg háromnak kell lenni.