

# Maximum likelihood

Rövid bevezetés

Medvegyev Péter

Corvinus

2011

## Maximum és kvázi 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 maximum 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  $f : \mathbb{R} \rightarrow \mathbb{R}$  konvex, akkor tetszőleges  $\xi$  változóra

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

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

Megjegyezzük, hogy konvex függvényekre mindig léteznek a  $\pm\infty$ -ben vett, esetlegesen végtelen értékű, határérték, így ha az  $\mathbf{E}(\xi)$  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  $\Omega$  valószínűségi mező minden  $\omega \in \Omega$  kimenetelére

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

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

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

# 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 expliciten feltettük, hogy az  $\mathbf{E}(\xi)$  és az  $\mathbf{E}(f(\xi))$  várható értékek léteznek.

### Theorem

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

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

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

**Bizonyítás:** A konvexitás miatt ismételten

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

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

$$f^-(t) \doteq \max(-f(t), 0) \leq |f(0)| + |f'_-(0)| \cdot |t|,$$

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

## Az elv indoklása

Legyen  $\mu$  tetszőleges mérték és legyenek  $p$  és  $q$  a  $\mu$  mérték szerinti sűrűségfüggvények. A  $\mu$  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

$$\int_{\Omega} p \cdot \ln q d\mu \leq \int_{\Omega} p \cdot \ln p d\mu, \quad (1)$$

feltéve, hogy mind a két oldal véges.

## Az elv indoklása

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

$$L(\xi_1, \dots, \xi_n, \theta) \doteq \prod_{k=1}^n f(\xi_k, \theta)$$

likelihood függvényt. Ha a  $\xi_k$  változók függetlenek, akkor a nagy számok törvénye alapján

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

ahol  $\xi$  tetszőleges olyan változó, amely eloszlása megegyezik a  $\xi_k$  változók közös eloszlásával.

## elv indoklása

Ha  $\theta_0$  a  $\theta$  paraméter valódi értéke, akkor a transzformált valószínűségi változók várható értékére vonatkozó formula alapján

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

Az (1) alapján

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

## Az elv indoklása

Ebből következően a  $\theta_0$  ismeretlen paraméter  $\hat{\theta}$  becslését úgy célszerű meghatározni, hogy maximalizáljuk a várható értéket közelítő

$$\frac{\ln L(\tilde{\xi}_1, \dots, \tilde{\xi}_n, \theta)}{n}$$

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

## Becslés hibája

Mekkora a becslés hibája, vagyis mekkora a  $\hat{\theta} - \theta_0$  eltérés?

Tekintsük az  $\ln L$  függvény Taylor-kifejtését a  $\theta_0$  helyes érték körül:

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

A kifejtést mint polinomot  $\theta$  szerint deriválva, felhasználva, hogy az első tag konstans

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

## Becsülés hibája

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

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

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

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

Ebből elemi átrendezéssel

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



## Becslés hibája

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

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

## Első tag

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

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

$$\eta_k \doteq \frac{d}{d\theta} \ln f(\xi_k, \theta_0)$$

változók közös várható értéke nulla.

## Első tag, várható érték

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

## Első tag, regularitási tulajdonság

Vegyük észre, hogy a számolás során kihasználtuk, hogy  $\theta_0$  a paraméter helyes értéke, így az  $\mathbf{E}$  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 ellenőrizni kell ezt a regularitási tulajdonságot.

## Első tag, szórás

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

$$\begin{aligned} & \frac{d}{d\theta} \int_{\mathbb{R}} \frac{d}{d\theta} \ln f(x, \theta) f(x, \theta) d\mu = \\ & = \frac{d}{d\theta} \int_{\mathbb{R}} \left( \frac{1}{f(x, \theta)} \frac{d}{d\theta} f(x, \theta) \right) f(x, \theta) d\mu = \\ & = \frac{d}{d\theta} \int_{\mathbb{R}} \frac{d}{d\theta} f(x, \theta) \frac{1}{f(x, \theta)} f(x, \theta) d\mu = \frac{d}{d\theta} \int_{\mathbb{R}} \frac{d}{d\theta} f(x, \theta) d\mu = \\ & = \frac{d^2}{d\theta^2} \int_{\mathbb{R}} f(x, \theta) d\mu = \frac{d^2}{d\theta^2} 1 = 0. \end{aligned}$$

## Első tag, szórás

Az integrálba bederiválva a szorzat deriválási szabálya szerint ugyanez

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

## 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  $\theta \doteq \theta_0$ , akkor az  $f(x, \theta_0)$  a valódi sűrűségfüggvény, így az integrálok a transzforált 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

$$\begin{aligned} \mathbf{E}(\eta^2) &= \mathbf{D}^2(\eta) = \mathbf{D}^2(\eta_k) \doteq \mathbf{D}^2 \left( \frac{d}{d\theta} \ln f(\xi_k, \theta_0) \right) = \\ &= \mathbf{E} \left( \left( \frac{d}{d\theta} \ln f(x, \theta_0) \right)^2 \right) = \\ &= -\mathbf{E} \left( \frac{d^2}{d\theta^2} \ln f(\xi_k, \theta_0) \right) \doteq \varphi. \end{aligned}$$



# Első tag, összefoglalás

Ebből a centrális határeloszlás tétel szerint

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

## 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

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

## Harmadik tag

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

$$\lim_{n \rightarrow \infty} \frac{T_3(\hat{\theta}_n - \theta_0)^2}{\sqrt{n}} = 0.$$

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

# Összefoglalás

Ilyenkor eloszlásban

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sqrt{n} (\hat{\theta} - \theta_0) = \\ = & \lim_{n \rightarrow \infty} = - \left( \frac{\frac{d \ln L}{d \theta} (\theta_0) + T_3 (\hat{\theta} - \theta_0)^2}{\sqrt{n}} \right) \left( \frac{\frac{d^2 \ln L}{d \theta^2} (\theta_0)}{n} \right)^{-1} = \\ & = -N(0, \sqrt{\varphi}) (-\varphi^{-1}) = N\left(0, \sqrt{\varphi^{-1}}\right). \end{aligned}$$

## A szórás becslése

Emlékeztetünk, hogy

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

A nagy számok törvénye szerint

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

## 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  $\hat{\theta}$ , 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  $\theta$  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  $\mathbf{E} \left( \ln \frac{d\mathbf{P}}{d\mathbf{Q}} \right)$ . Érdeemes hangsúlyozni, hogy az induló feltételek szerint a sűrűségfüggvények léteznek. és ugyancsak impliciten 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

$$I(\mathbf{P} | \mathbf{Q}) \stackrel{\circ}{=} \mathbf{E} \left( \ln \frac{d\mathbf{P}}{d\mathbf{Q}} \right).$$

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

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

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 akár csak a klasszikus esetben konzisztens becslése annak a  $\theta_*$ -nak, amely minimalizálja a KL-távolságot. Ezenban 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

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

Ha a  $\theta$  paraméterter több dimenziós, akkor ez

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

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  $\Phi$  pozitív semidefinite.

## A hiba becslése

Vezessük be az

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

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  $\theta_*$  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

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

A két érték azonban nem azonos.

## A normalitás esetén való példa

Példaként tekintsük a normális eloszlás paramétereinek becslését.  
Ekkor

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

## A várható érték és szórás becslése

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

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

ahol

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

# A várható érték és szórás becslése

Ugyanis a számláló

$$\begin{aligned} & \sum_k (\xi_k - \bar{\xi}_n)^2 + n (\bar{\xi}_n - \mu)^2 = \\ &= \sum_k \xi_k^2 - 2 \sum_k \xi_k \bar{\xi}_n + n \bar{\xi}_n^2 + n (\bar{\xi}_n - \mu)^2 = \\ &= \sum_k \xi_k^2 - 2 \sum_k \xi_k \bar{\xi}_n + n \bar{\xi}_n^2 + n \bar{\xi}_n^2 + n \mu^2 - 2n \bar{\xi}_n \mu = \\ &= \sum_k \xi_k^2 - 2n \bar{\xi}_n + n \bar{\xi}_n^2 + n \bar{\xi}_n^2 + n \mu^2 - 2n \bar{\xi}_n \mu = \\ &= \sum_k \xi_k^2 + n \mu^2 - 2n \bar{\xi}_n \mu = \sum_k (\xi_k - \mu)^2. \end{aligned}$$

# A várható érték és szórás becslése

Az első egyenletből

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

A  $\mu$  becslését a másodikba betéve

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

A  $\sigma$ -val beszorozva és átrendezve

$$\hat{\sigma}_n^2 = \frac{\sum_k (\xi_k - \bar{\xi}_n)^2}{n} \rightarrow \mathbf{D}^2(\xi) = \sigma^2.$$



## A két mátrix kiszámolása

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

$$\begin{aligned} A(\mu, \sigma^2) &\doteq \mathbf{E} \left( \begin{pmatrix} \partial^2 I / \partial \mu^2 & \partial^2 I / \partial \mu \partial V \\ \partial^2 I / \partial \mu \partial V & \partial^2 I / \partial V^2 \end{pmatrix} \right) = \\ &= \mathbf{E} \begin{pmatrix} -1/\sigma^2 & -(\xi - \mu) / \sigma^4 \\ -(\xi - \mu) / \sigma^4 & 1 / (2\sigma^4) - 2 \frac{(\xi_k - \mu)^2}{2\sigma^6} \end{pmatrix} = \\ &= \begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & 1 / (2\sigma^4) - \frac{\sigma^2}{\sigma^6} \end{pmatrix} = \begin{pmatrix} -1/\sigma^2 & 0 \\ 0 & -1/2\sigma^4 \end{pmatrix} = . \end{aligned}$$

# A két mátrix kiszámolása

Az elsőrendű parciális deriváltak

$$\left( \frac{\partial l}{\partial \mu} \quad \frac{\partial l}{\partial \sigma^2} \right) = \left( \frac{x-\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \right)$$

A diadikus szorzat várható értéke

$$\begin{aligned} B(\mu, \sigma^2) &\doteq \mathbf{E} \left( \begin{array}{cc} \frac{(\xi-\mu)^2}{\sigma^4} & -\frac{\xi-\mu}{\sigma^4} + \frac{(\xi-\mu)^3}{2\sigma^6} \\ -\frac{\xi-\mu}{\sigma^4} + \frac{(\xi-\mu)^3}{2\sigma^6} & \left( \frac{1}{2\sigma^2} - \frac{(\xi-\mu)^2}{2\sigma^4} \right)^2 \end{array} \right) = \\ &= \mathbf{E} \left( \begin{array}{cc} 1/\sigma^2 & \frac{(\xi-\mu)^3}{2\sigma^6} \\ \frac{(\xi-\mu)^3}{2\sigma^6} & \left( \frac{1}{2\sigma^2} - \frac{(\xi-\mu)^2}{2\sigma^4} \right)^2 \end{array} \right). \end{aligned}$$

# Ferdeség és kurtózis

Ha

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

a ferdeségi mutató és

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

a kurtózis akkor

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

## A két mátrix kiszámolása

ugyanis

$$\begin{aligned}\mathbf{E} \left( \left( -\frac{1}{2\sigma^2} + \frac{(\xi - \mu)^2}{2\sigma^4} \right)^2 \right) &= \frac{1}{4\sigma^4} \mathbf{E} \left( \left( -1 + \left( \frac{\xi - \mu}{\sigma} \right)^2 \right)^2 \right) = \\ &= \frac{1}{4\sigma^4} \left( 1 - 2\mathbf{E} \left( \frac{\xi - \mu}{\sigma} \right)^2 + \mathbf{E} \left( \frac{\xi - \mu}{\sigma} \right)^4 \right) \doteq \\ &\doteq \frac{1}{4\sigma^4} (1 - 2 + \kappa) = \frac{1}{4\sigma^4} (\kappa - 1).\end{aligned}$$

## 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  $C(\mu, \sigma^2)$  mátrixot.

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

Így

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

## 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

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

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.