Proposition 1 (Jensen-egyenlőtlenség) *Ha $f : \mathbb{R} \to \mathbb{R}$ konvex, akkor tetszőleges ξ változóra*

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

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

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

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

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

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[1]. De ekkor az alapul vett Ω 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

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

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

3. Tegyük fel, hogy $\mathbf{M}(\xi) = \infty$. 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

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

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

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

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

5. Végezetül ha $\mathbf{M}(f(\xi)) = -\infty$, és van olyan t, amelyre $f'_{-}(t) > 0$, akkor az imént említett módon $\mathbf{M}(\xi) = -\infty$, és az egyenlőtlenség a már bemutatott indokolható. Ha $f'_{-} \leq 0$, és van olyan t, hogy $f'_{-}(t) < 0$, akkor egyrészt az

[1] 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.

f csökken, másrészt a már korábban elmondottak miatt $\mathbf{M}(\xi) = \infty$, vagyis $f(\mathbf{M}(\xi)) \leq f(\xi)$. Ha $f'_{-} = 0$, akkor az f véges konstans, és az $\mathbf{M}(f(\xi)) = -\infty$ nem teljesülhet.

□

Az egyenlőtlenségben explicite feltettük, hogy az $\mathbf{M}(\xi)$ és az $\mathbf{M}(f(\xi))$ várható értékek léteznek. Éppen ezért nem érdektelen a következő észrevétel:

Proposition 2 *Ha $f : \mathbb{R} \to \mathbb{R}$ konvex és az $\mathbf{M}(\xi)$ várható érték véges, akkor*

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

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

Bizonyítás: A konvexitás miatt ismételten

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

amiből

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

következésképpen

$$f^{-}(t) \mathrel{\hat{=}} \max(-f(t), 0) \leq |f(0)| + \left|f^{-}_{-}(0)\right| \cdot |t|,$$

vagyis

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

□

Example 3 *A maximum likelihood elv indoklása.*

Legyen μ tetszőleges mérték és legyenek p és q a μ mérték szerinti sűrűségfüggvények[2]. A logaritmus függvény konkáv, így a Jensen-egyenlőtlenség szerint

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

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 egyenlőtlenség alkalmazásaként tegyük fel, hogy a $\xi_1, \xi_2, \ldots, \xi_n$ változók azonos eloszlásúak, és a közös sűrűségfüggvényük

[2] A μ lehet például a számláló mérték. Ilyenkor a diszkrét eloszlásokat kezeljük és az integrálok egyszerű összegek.

legyen $f(x,\theta)$, ahol θ egy ismeretlen paraméter amely értékét a $(\xi_k)_{k=1}^n$ változók megfigyelése alapján próbáljuk megbecsülni. Képezzük az

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

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

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

ahol ξ tetszőleges olyan változó, amely eloszlása megegyezik a ξ_k változók közös eloszlásával[3]. Ha θ_0 a θ paraméter valódi értéke, akkor a transzformált valószínűségi változók várható értékére vonatkozó formula alapján

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

Az (1) alapján

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

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

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

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

$\square$

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:

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

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

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

[3]Feltéve persze, hogy a várható érték létezik.

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

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

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

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

Ebből elemi átrendezéssel

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

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

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

Vizsgáljuk meg az egyes kifejezések határértékét!

1.

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

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

$$\eta_k\mathrel{\hat{=}}\frac{d}{d\theta}\ln f(\xi_k,\theta_0)$$

változók közös várható értékét.

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

Vegyük észre, hogy a számolás során kihasználtuk, hogy θ_0 a paraméter helyes értéke, így az $\mathbf{M}$ 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. 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

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

Az integrálba bederiválva a szorzat deriválási szabálya szerint ugyanez

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

Ahol ismételten többször kihasználtuk, hogy az $f(x,\theta)$ kifejezés elég jó ahhoz, hogy a deriválás és az integrálás felcserélhető. Ebből a már többször használt

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

azonosság szerint[4]

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

[4]Ugyanis az összeg nulla.

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

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

vagyis

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

Ebből a centrális határeloszlás tétel szerint

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

2. Ugyanakkor a nagy számok törvénye szerint

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

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

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

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.

Ilyenkor eloszlásban

$$\lim_{n\to\infty}\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = -N(0,\sqrt{\varphi})\left(-\varphi^{-1}\right) = N\left(0,\sqrt{\varphi^{-1}}\right).$$

Emlékeztetünk, hogy

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

A nagy számok törvénye szerint

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