Statisztikai becslések jósága.

Ebben a jegyzetben azzal a kérdéssel foglalkozunk, hogy bizonyos természetes tulajdonságokat teljesítő statisztikai modellekben milyen jó paraméterbecslést lehet adni.

Annak érdekében, hogy jobban megértsük milyen eredményt várhatunk, tekintsük a következő példát. Vegyünk egy $(X_1, \dots, X_n)$ mintát, amely független, egyforma eloszlású valószínűségi változókból áll, és becsüljük meg az X_j valószínűségi változók $\mu = EX_j$ várható értékét. A természetes becslés az $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ kifejezés. Ez a μ paraméter torzítatlan becslése, amelynek hibája a centrális határeloszlástétel szerint közel normális eloszlású, nulla vérható értékkel és $\frac{\operatorname{Var} X_1}{n}$ szórásnégyzettel. Felmerül a kérdés, hogy lehet-e jobb becslést adni, illetve lehet-e hasonlóan jó becslést adni általánosabb esetben is.

Be fogjuk látni, hogy bizonyos regularitási feltételek teljesülése esetén nem lehet jobb becslést adni, mint amit a fenti példa sugall. Alsó becslést adunk egy paraméterbecslés hibájának a szórásnégyzetére, és a hiba nagyságrendje egy n elemű minta esetén $O(\frac{1}{n})$ nagyságrendű lesz. A szórásnégyzet hibájára adott alsó becslésben megjelenik egy fontos paraméter, az úgynevezett Fisher információ.

Másik irányban be fogjuk látni, hogy (még több regularitási feltétel teljesülése esetén) létezik olyan paraméter becslés, az úgynevezett maximum likelihood becslés, amely olyan jó becslést ad (aszimptotikusan normális eloszlású hibával), mint amilyet a fenti példa sugall. Ezen becslés hibájának a megadásában is megjelenik a Fisher információ.

Felmerül még az a kérdés, hogy az ezekben az eredményekben szereplő regularitási feltételek valóban fontosak-e. Fogunk példát mutatni olyan a regularitási feltételt nem teljesitő természetes modellre, amelyben lényegesen jobb paraméterbecslést lehet csinálni, mint amit a fent emlitett eredmények sugallnak.

Az első kérdés, amivel foglalkozni fogunk a következő. Legyen adva egy X valószínűségi változó egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, amely értékeit az R^n Euklideszi térben veszi fel, eloszlása valamilyen P_ϑ mérték, ahol $\vartheta \in \Theta$, és a Θ paraméterhalmaz a számegyenes egy nyílt részhalmaza. Feltesszük továbbá, hogy a P_ϑ mértéknek van egy $f_\vartheta(x)$, $\vartheta \in \Theta$, $x \in R^n$, sűrűségfüggvénye egy σ-véges ν domináló mérték szerint az R^n Euklideszi téren. Milyen jó becslést tudunk adni a ϑ paraméter valamely $\psi(\vartheta)$ függvényére? Először egy alsó becslést kívánunk adni a $\psi(\vartheta)$ függvény torzítatlan becsléseinek a hibájára, azaz olyan $T(X)$ becslések hibájára, amelyekre

$$E_\vartheta T(X)=\int T(x)f_\vartheta(x)\nu(\,dx)=\psi(\vartheta),$$

ha X eloszlásfüggvénye P_ϑ. Bebizonyítjuk ilyen becslésekre a Cramer–Rao egyenlőtlenséget, amely alkalmas feltételek teljesülése esetén alsó becslést ad a $\mathrm{Var}_\vartheta T(X)=E_\vartheta T^2(X)-\psi^2(\vartheta)$ szórásnégyzetre. Külön érdekel minket az a speciális eset, amikor az X valószínűségi változó $X=(X_1,\dots,X_n)$ alakú, ahol $X_1,\dots,X_n$ független, egyforma eloszlású valószínűségi változók sorozata, valamely $P_\vartheta^{(0)}$, $\vartheta\in\Theta$, eloszlásfüggvénnyel.

E jegyzetben főleg azt az esetet vizsgáljuk, amikor a Θ paramétertartomány az egydimenziós Euklideszi tér nyílt részhalmaza, de röviden ismertetni fogom ezen eredmény általánosítását arra az esetre, amikor a Θ paramétertartomány a k-dimenziós R^k Euklideszi tér nyílt részhalmaza.

Egy másik e jegyzetben tárgyalt probléma a maximum likelihood becslés viselkedéséről szól. Heurisztikus szinten elmagyarázom, hogy alkalmas feltételek teljesülése esetén a maximum likelihood becslés aszimptotikusan optimális. Majdnem olyan jó becslést ad, mint amilyet a Cramer–Rao egyenlőtlenség még megenged. Azután megfogalmazok egy pontos tételt a maximum likelihood becslés tulajdonságairól, de a bizonyítás részleteit elhagyom.

A Cramer–Rao egyenlőtlenség megfogalmazásában felteszem, hogy úgynevezett reguláris esettel fogunk foglalkozni, ahol az $f_\vartheta(x)$ sűrűségfüggvény elég sima függvénye az x és ϑ változóknak, valamint bizonyos integrálokban a ϑ változó szerinti deriválás és az x változó szerinti integrálás felcserélhető. Pontosabban fogalmazva, felteszem, hogy bizonyos $U(x)$ függvényekre

$$\frac{\partial^j}{\partial\vartheta^j}\int U(x)f_\vartheta(x)\nu(\,dx)=\int U(x)\frac{\partial^j}{\partial\vartheta^j}f_\vartheta(x)\nu(\,dx),\quad j=1,2.$$

Megadom azokat az eseteket, amelyekben ezt a formulát használni fogjuk a bizonyításokban. Egyrészt

$$\frac{\partial^j}{\partial\vartheta^j}\int f_\vartheta(x)\nu(\,dx)=\int\frac{\partial^j}{\partial\vartheta^j}f_\vartheta(x)\nu(\,dx),\quad j=1,2,$$

másrészt

$$\frac{\partial}{\partial\vartheta}\int T(x)f_\vartheta(x)\nu(\,dx)=\int\frac{\partial}{\partial\vartheta}T(x)f_\vartheta(x)\nu(\,dx),$$

ahol $T(x)$ a vizsgált becslőfüggvény. Megjegyzem, hogy $\int f_\vartheta(x)\nu(\,dx) = 1$, és $\int T(x) f_\vartheta(x)\nu(\,dx) = E_\vartheta T(x)$.

Bevezetem a Fisher információ fogalmát, amely megjelenik a Cramer–Rao egyenlőtlenség megfogalmazásában.

Fisher információ definiciója. *Legyen adva egy σ-véges ν domináló mérték szerinti $f_\vartheta(x)$, $x \in R^n$, sűrűségfügvényeknek egy $\vartheta \in \Theta$, paraméterrel indexelt családja, mely sűrűségfüggvények elég simán függnek az x és ϑ változóktól. E függvénycsalád Fisher információja egy $\vartheta \in \Theta$ pontban*

$$I(\vartheta) = E_\vartheta \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(x) \right)^2 = \int \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(x) \right)^2 f_\vartheta(x)\nu(\,dx). \tag{1}$$

Megjegyzés. Az (1) formulában, és hasonló képletekben a $0 \cdot \infty = 0$ konvenciót fogjuk alkalmazni, azaz definició szerint $\int_{\{x\colon f_\vartheta(x)=0\}} S(x) f_\vartheta(x)\nu(\,dx) = 0$ egy $S(x)$ függvény integráljára. Ez a formula akkor is érvényes, ha $S(x) = \pm\infty$ bizonyos pontokban. A reguláris viselkedés feltételében szereplő azonosságokban szereplő integrálok mindkét oldalán az $\{x\colon\ f_\vartheta(x) > 0\}$ halmazon integrálunk.

Megfogalmazom a Cramer–Rao egyenlőtlenséget.

Cramer–Rao egyenlőtlenség. *Legyen adva egy X valószínűségi változó egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, amely értékeit az R^n Euklideszi térben veszi fel, eloszlása valamely P_ϑ, $\vartheta \in \Theta$, mértékcsaládba tartozik, és mindegyik P_ϑ, $\vartheta \in \Theta$, mértéknek létezik $f_\vartheta(x)$, $x \in R^n$, $\vartheta \in \Theta$, sűrűségfüggvénye egy σ-véges ν domináló mérték szerint az R^n Euklideszi téren. Továbbá a Θ paraméterhalmaz a számegyenes egy nyílt részhalmaza. Feltesszük, hogy az f_ϑ sűrűségfüggvények regulárisan viselkednek. Legyen adva egy $\psi(\vartheta)$ függvény, ahol $\psi(\cdot)$ differenciálható függvény a Θ halmazon, és annak egy torzítatlan $T(X)$ becslése, azaz legyen $E_\vartheta T(X) = \psi(\vartheta)$ minden $\vartheta \in \Theta$ paraméterre. Ekkor*

$$Var(T(X)) \ge \frac{|\psi'(\vartheta)|^2}{I(\vartheta)} \tag{2}$$

abban az esetben, ha az X valószínűségi változó eloszlása P_ϑ. A (2) képletben szereplő $I(\vartheta)$ kifejezés az $f_\vartheta(x)$, $\vartheta \in \Theta$, sűrűségfüggvénycsalád Fisher információja a ν domináló mérték szerint a ϑ pontban.

Tekintsük azt a speciális esetet, amikor $X=(X_1,\dots,X_n)$, és $X_1,\dots,X_n$ függgetlen, egyforma eloszlású valószínűségi változók valamilyen $P_\vartheta^{(0)}$, $\vartheta\in\Theta$, eloszlással a számegyenesen, amelynek létezik $f_\vartheta^{(0)}(x)$, $\vartheta\in\Theta$, sűrűségfüggvénye egy ν σ-véges mérték szerint a számegyenesen. Legyen

$$I_1(\vartheta)=E_\vartheta\left(\frac{\partial}{\partial\theta}\log f_\vartheta^{(0)}(x)\right)^2=\int\left(\frac{\partial}{\partial\vartheta}\log f_\vartheta^{(0)}(x)\right)^2 f_\vartheta^0(x)\nu(\,dx). \tag{3}$$

Ekkor $I(\vartheta)=nI_1(\vartheta)$, ahol az $I(\vartheta)$ Fisher információ kiszámolásában a $\nu(dx_1)\times\cdots\times\nu(dx_n)$ szorzatmérték a domináló mérték.

Ha $T(X)=T(X_1,\dots,X_n)$ a $\psi(\vartheta)$ függvény torzítatlan becslése, és az X_1 valószínűségi változó eloszlása $P_\vartheta^{(0)}$, akkor

$$\mathit{Var}\,(T(X_1,\dots,X_n))=\mathit{Var}\,(T(X))\ge\frac{|\psi'(\vartheta)|^2}{nI_1(\vartheta)}. \tag{4}$$

A Cramer–Rao egyenlőtlenség bizonyítása. A (2) formulát az alábbi két U és V valószínűségi változóra vonatkozó egyenlőtlenség segítségével fogjuk bizonyítani.

$$|\mathrm{Cov}\,(U,V)|^2\le\mathrm{Var}\,U\cdot\mathrm{Var}\,V. \tag{5}$$

Ez az egyenlőtlenség következik a Cauchy–Schwarz egyenlőtlenségből, mert

$$\begin{aligned}|\mathrm{Cov}\,(U,V)|^2 &= [E(U-EU)(V-EV)]^2\\ &\le E(U-EU)^2E(V-EV)^2=\mathrm{Var}\,U\cdot\mathrm{Var}\,V.\end{aligned}$$

Az (5) egyenlőtlenséget $U(x)=\frac{\partial}{\partial\vartheta}\log f_\vartheta(x)$ és $V(x)=T(x)$ választással fogjuk alkalmazni az $(R^n,\mathcal{B}^n,P_\vartheta)$ valószínűségi mezőn a

$$P_\vartheta(\,dx)=f_\vartheta(x)\nu(\,dx)$$

valószínűségi mértékkel. A bizonyítás érdekében kiszámoljuk az $\mathrm{Var}\,V(x)$ szórásnégyzetet és $\mathrm{Cov}\,(U(x),V(x))$ kovarianciát. Először azt látjuk be, hogy

$$E_\vartheta U(x)=0,$$

mert

$$\begin{aligned}E_\vartheta U(x)=\int\frac{\partial}{\partial\vartheta}(\log f_\vartheta(x))f_\vartheta(x)\nu(\,dx) &= \int\frac{\partial}{\partial\vartheta}f_\vartheta(x))\nu(\,dx)\\ &= \frac{\partial}{\partial\vartheta}\int f_\vartheta(x))\nu(\,dx)=0,\end{aligned}$$

E számolásban kihasználtuk, hogy az integrálás és deriválás sorrendje felcserélhető a regularitási feltétel miatt, valamint

$$\int f_\vartheta(x)\nu(\,dx) = 1 \quad \text{minden } \vartheta \in \Theta \text{ paraméterre,}$$

ahonnan

$$\frac{\partial}{\partial\vartheta}\int f_\vartheta(x)\nu(\,dx) = 0.$$

Innen $\operatorname{Var} U(x) = E_\vartheta U(x)^2$, $\operatorname{Cov}(U(x), V(x)) = E_\vartheta U(x)V(x)$, ezért

$$\operatorname{Var} U(x) = E_\vartheta U(x)^2 = \int \left(\frac{\partial}{\partial\vartheta}(\log f_\vartheta(x))\right)^2 f_\vartheta(x)\nu(\,dx) = I(\vartheta)$$

és

$$\begin{aligned}
\operatorname{Cov}(U(x), V(x)) &= ET(x)U(x) = \int T(x)\frac{\partial}{\partial\vartheta}(\log f_\vartheta(x)) f_\vartheta(x)\nu(\,dx) \\
&= \int T(x)\frac{\partial}{\partial\vartheta} f_\vartheta(x))\nu(\,dx) \\
&= \frac{\partial}{\partial\vartheta}\int T(x) f_\vartheta(x))\nu(\,dx) = \psi'(\vartheta).
\end{aligned}$$

Itt megint kihasználtuk, hogy a regularitási tulajdonság miatt az integrálás és deriválás sorrendje felcserélhető, valamint azt, hogy $\int T(x)f_\vartheta(x))\nu(\,dx) = E_\vartheta T(x) = \psi(\vartheta)$. Ezért az (5) formula szerint

$$\psi'(\vartheta)^2 \le \operatorname{Var} T(x) I(\vartheta).$$

Innen következik a (2) formula, mert a $T(X)$ valószínűségi változó szórásnégyzete az $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn megyezik a $T(x)$ valószínűségi változó szórásnégyzetével az $(R^n, \mathcal{B}^n, P_\vartheta)$ valószínűségi mezőn.

Be akarjuk bizonyítani az $I(\vartheta) = nI_1(\vartheta)$ azonosságot, ahol az $I_1(\vartheta)$ kifejezést a (3) relációban definiáltuk. Ennek érdekében először megadjuk az (1) formulában definiált $I(\vartheta)$ Fisher információnak egy másik definicióját. (Feltesszük, hogy az $f_\vartheta(x)$ sűrűségfüggvénynek a ϑ paraméter szerint van második deriváltja is, és az is szépen viselkedik.) Megmutatjuk, hogy

$$I(\vartheta) = -\int \frac{\partial^2}{\partial\vartheta^2}(\log f_\vartheta(x)) f_\vartheta(x)\nu(\,dx). \tag{6}$$

Valóban,

$$\frac{\partial^2}{\partial\vartheta^2}\log f_\vartheta(x)=\frac{\frac{\partial^2}{\partial\vartheta^2}f_\vartheta(x)}{f_\vartheta(x)}-\left(\frac{\frac{\partial}{\partial\vartheta}f_\vartheta(x)}{f_\vartheta(x)}\right)^2,$$

és a jobboldal első tagjának az integrálja eltűnik, azaz

$$\int\frac{\frac{\partial^2}{\partial\vartheta^2}f_\vartheta(x)}{f_\vartheta(x)}f_\vartheta(x)\nu(\,dx)=\frac{\partial^2}{\partial\vartheta^2}\int f_\vartheta(x)\nu(\,dx)=0.$$

Ebben a számolásban megint kihasználtuk, hogy a regularitási feltételeink megengedik, hogy bizonyos integrálások és deriválások sorrendjét felcseréljük.

Ezt felhasználva, és az előző azonosságot kiintegrálva megkapjuk az

$$\begin{aligned}-\int\frac{\partial^2}{\partial\vartheta^2}(\log f_\vartheta(x))f_\vartheta(x)\nu(\,dx) &= \int\left(\frac{\frac{\partial}{\partial\vartheta}f_\vartheta(x)}{f_\vartheta(x)}\right)^2 f_\vartheta(x)\nu(\,dx)\\ &= \int\left(\frac{\partial}{\partial\vartheta}\log f_\vartheta(x)\right)^2 f_\vartheta(x)\nu(\,dx)=I(\vartheta)\end{aligned}$$

azonosságot, ami azt jelenti, hogy a (6) azonosság igaz.

A (6) azonosság segítségével a következőképp kapjuk meg az $I(\vartheta)=nI_1(\vartheta)$ azonosságot.

Mivel $X_1,\dots,X_n$ független és egyforma eloszlású valószínűségi változók, ezért ezek együttes sűrűségfüggénye $f_\vartheta(x)=f_\vartheta^{(0)}(x_1)\cdots f_\vartheta^{(0)}(x_n)$ az $x=(x_1,\dots,x_n)$ pontban a $\nu(\,dx_1)\times\cdots\times\nu(\,dx_n)$ szorzatmérték szerint. Ezért az $X=(X_1,\dots,X_n)$ vektor Fisher információja a (6) képlet szerint

$$\begin{aligned}I(\vartheta) &= -\int\frac{\partial^2}{\partial\vartheta^2}\log\left(\prod_{k=1}^n f_\vartheta^{(0)}(x_k)\right)\prod_{k=1}^n f_\vartheta^{(0)}(x_k)\nu(\,dx_k)\\ &= -n\int\frac{\partial^2}{\partial\vartheta^2}(\log f_\vartheta^{(0)}(x_1))f_\vartheta^{(0)}(x_1)\nu(\,dx_1).\end{aligned}$$

A (6) formula bizonyításához hasonlóan kapjuk, hogy

$$I_1(\vartheta)=-\int\frac{\partial^2}{\partial\vartheta^2}(\log f_\vartheta^{(0)}(x_1))f_\vartheta^{(0)}(x_1)\nu(\,dx_1)$$

a (3) formulában definiált $I_1(\vartheta)$ függvényre. (Valójában ez a képlet a (6) formula speciális esete.) Ezért az $I(\vartheta)=nI_1(\vartheta)$ azonosság érvényes. A (4) formula következik ebből az azonosságból, ha a (2) formulát az $X=$

$(X_1, \dots, X_n)$ vektor sűrűségfüggvényére alkalmazzuk. A Cramer–Rao egyenlőtlenséget bebizonyítottuk.

Röviden ismertetem a Cramer–Rao egyenlőtlenséget abban az általánosabb esetben, ha a Θ paramétertartomány az R^k Euklideszi tér nyílt részhalmaza valamely $k \geq 1$ paraméterrel. A bizonyítás azon részleteit, amelyek hasonlóak az egydimenziós esethez elhagyom. Egyedül a bizonyítás azon lépéséről fogok írni, ahol új, az egydimenziós eset tárgyalásában nem szereplő érvelést alkalmazunk.

E problémában az X valószínűségi változó eloszlása valamely P_ϑ mérték egy $f_\vartheta(x)$ sűrűségfüggvénnyel az R^n téren egy ν σ-véges domináló mértékkel, és a ϑ paraméter a $\vartheta \in \Theta$ feltételt egy k-dimenziós Θ nyílt paraméterhalmazzal teljesíti.

Először definiálni kell az $I(\vartheta)$ Fisher információt egy $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Theta$ pontban. Ez az az $(A_{i,j}(\vartheta))$, $1 \leq i, j \leq k$, $k \times k$ szimmetrikus (pozitív szemidefinit) mátrix, amelynek elemei az

$$A_{i,j}(\vartheta) = \int \frac{\partial}{\partial \vartheta_i} (\log f_\vartheta(x)) \frac{\partial}{\partial \vartheta_j} (\log f_\vartheta(x)) f_\vartheta(x) \nu(\,dx), \quad 1 \leq i, j \leq k,$$

számok.

Legyen $T(X)$ torzítatlan becslése egy $\psi(\vartheta)$ függvénynek. Ekkor az egydimenziós esethez hasonló feltételek mellett

$$\operatorname{Var}(T(X)) \geq \left(\frac{\partial}{\partial \vartheta_1} \psi(\vartheta), \dots, \frac{\partial}{\partial \vartheta_k} \psi(\vartheta) \right) I(\vartheta)^{-1} \left(\frac{\partial}{\partial \vartheta_1} \psi(\vartheta), \dots, \frac{\partial}{\partial \vartheta_k} \psi(\vartheta) \right)^* \tag{7}$$

minden $\vartheta \in \Theta$ pontban, ahol * egy vektor transzponáltját jelöli.

Az egydimenziós esethez hasonlóan az $I(\vartheta)$ Fisher információt megadó mátrix elemeit deiniálhatjuk a következő módon is.

$$A_{i,j}(\vartheta) = -\int \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} (\log f_\vartheta(x)) f_\vartheta(x) \nu(\,dx), \quad 1 \leq i, j \leq k.$$

Ezt felhasználva megmutathatjuk, hogy független, azonos eloszlású koordinátákból álló vektorok becslésére a (4) formulához hasonló becslést lehet megadni vektor értékű paraméter tartomány esetében is.

Legyen $X = (X_1, \dots, X_n)$, ahol $X_1, \dots, X_n$ független, azonos eloszlású valószínűségi változók sorozata, amelyeknek $P_\vartheta^{(0)}$, $\vartheta \in \Theta$, eloszlásának van $f_\vartheta^{(0)}(x)$, sűrűségfüggvénye egy ν σ-véges mérték szerint. Ki lehet számolni az $f^{(0)}(\vartheta)$ sűrűségfüggvényhez és ν domináló mértékhez tartozó $I_1(\vartheta)$ Fisher információ mátrixát, és azt kapjuk, hogy $I(\vartheta) = nI_1(\vartheta)$, ahol $I(\vartheta)$ az X valószínűségi vektor sűrűségfüggvényéhez és a hozzá tartozó domináló mértékhez tartozó Fisher információ mátrixa. Innen következik, hogy ha az $X = (X_1, \dots, X_n)$ véletlen vektor teljesíti a Cramer–Rao egyenlőtlenség feltételeit, akkor érvényes rá a (7) becslés azon változata, amelyben az $I(\vartheta)$ Fisher információ mátrixot az $nI_1(\vartheta)$ mátrixszal helyettesítjük.

A Cramer–Rao egyenlőtlenség bizonyítása vektor értékű Θ paramétertartomány esetében hasonló az egydimenziós paramétertartomány esetéhez. Ezért a bizonyításnak csak azt a pontját vázolom, ahol új módszer szükséges. Ez a (7) formula bizonyítása.

A bizonyítás első lépésében az egydimenziós esethez hasonlóan érvelünk. Az (5) formula segítségével hasonló becsléseket teszünk, mint az egydimenziós esetben. Jelen esetben is $V(x) = T(x)$ függvényt választunk. Az $U(x)$ függvényt az $U_j(x) = \frac{\partial}{\partial \vartheta_j}(\log f_\vartheta(x))$, $j = 1, \dots, k$, függvények és tetszőleges $c_1, \dots, c_k$ konstansok segítségével az $U(x) = \sum_{j=1}^k c_j U_j(x)$ képlettel definiáljuk. Ilyen választással be tudjuk látni az (5) egyenlőtlenség felhasználásával, hogy

$$\left| \sum_{j=1}^k c_j \frac{\partial}{\partial \vartheta_j} \psi(\vartheta) \right|^2 \le \operatorname{Var} T(x) \left(\sum_{i=1}^k \sum_{j=1}^k c_i c_j A_{i,j}(\vartheta) \right), \tag{8}$$

ahol az $A_{i,j}(\vartheta)$ együtthatók a $I(\vartheta)$ Fisher mátrix definiciójában szereplő számok, és $c_1, \dots, c_k$ tetszőleges konstansok. Ezt az egyenlőtlenséget számunkra hasznosabb alakban fogjuk felírni.

Legyen $e_1, \dots, e_k$, az $(A_{i,j}(\vartheta))$, $1 \le i, j \le k$, Fisher információs mátrix ortonormált sajátvektoraiból álló bázis az R^k Euklideszi térben, és legyenek $\lambda_1, \dots, \lambda_k$ a hozzájuk tartozó sajátértékek. Fejezzük ki a (8), illetve a bizonyítandó (7) egyenlőtlenségeket az $e_1, \dots, e_k$, koordinátarendszerben. Ennek érdekében írjuk az

$$S(\vartheta) = \left(\frac{\partial}{\partial \vartheta_1} \psi(\vartheta), \dots, \frac{\partial}{\partial \vartheta_k} \psi(\vartheta) \right)$$

vektort $S(\vartheta) = \sum_{j=1}^k \alpha_j e_j$ alakban alkalmas (a ϑ paramétertől függő) $\alpha_1, \dots, \alpha_k$ együtthatókkal. Akkor felhasználva, hogy a (8) formula tetszőleges $c_1, \dots, c_k$ együtthatókra érvényes, az új koordinátarendszerben azt kapjuk, hogy

$$\left(\sum_{j=1}^k c_j \alpha_j \right)^2 \le \operatorname{Var} T(x) \left(\sum_{j=1} \lambda_j c_j^2 \right)$$

tetszőleges $c_1, \dots, c_k$ konstansokkal. A bizonyítandó (7) egyenlőtlenség pedig

$$\operatorname{Var} T(x) \ge \sum_{j=1}^{k} \frac{\alpha_j^2}{\lambda_j}$$

alakban írható. Ezt a relációt megkapjuk az előző egyenlőtlenségből, ha azt $c_j = \frac{\alpha_j}{\lambda_j}$ választással alkalmazzuk. Ez ugyanis azt adja, hogy

$$\left(\sum_{j=1}^{k} \frac{\alpha_j^2}{\lambda_j}\right)^2 \le \operatorname{Var} T(x) \left(\sum_{j=1}^{k} \frac{\alpha_j^2}{\lambda_j}\right).$$

A Cramer–Rao egyenlőtlenség tárgyalásának a végén teszek egy megjegyzést arról, hogy milyen alsó becslést lehet adni egy nem feltétlenül torzítatlan becslés hibájára, ha a Cramer–Rao egyenlőtlenség többi feltétele teljesül. (Most egydimenziós paraméterrel rendelkező modelleket vizsgálunk.)

Tekintsük egy $\psi(\vartheta)$ függvény nem feltétlenül torzítatlan $T(X)$ becslését, és legyen

$$b_T(\vartheta) = E_\vartheta(T(x)) - \psi(\vartheta).$$

Ha a Cramer–Rao egyenlőtlenség egyéb feltételei teljesülnek, akkor felírhatjuk, hogy

$$\operatorname{Var}(T(x)) \ge \frac{(\psi'(\vartheta) + b'_T(\vartheta))^2}{I(\vartheta)} \quad \text{minden } \vartheta \in \Theta \text{ paraméterre}$$

a Cramer–Rao egyenlőtlenség alapján, ha azt a $\psi^*(\vartheta) = E_\vartheta T(x) = \psi(\vartheta) + b_T(\vartheta)$ függvényre alkalmazzuk.

Vezessük be az $R_\vartheta(T) = E_\vartheta(T(x) - \psi(\vartheta))^2$ mennyiséget, amit úgy hívnak, hogy a $\psi(\vartheta)$ függvény $T(X)$ becslésének a négyzetes rizikója, és ami jól méri a $T(X)$ becslés hibáját. Ekkor felírhatjuk a Steiner azonosság alapján, hogy

$$R_\vartheta(T) = \operatorname{Var}(T(x)) + b_T^2(\vartheta),$$

és

$$R_\vartheta(T) \ge \frac{(\psi'(\vartheta) + b'_T(\vartheta))^2}{I(\vartheta)} + b_T^2(\vartheta) \quad \text{minden } \vartheta \in \Theta \text{ paraméterre.}$$

Megjegyzem, előfordulhat, hogy torzított becsléssel kisebb négyzetes rizikó érhető el, mint a lehető legjobb torzítatlan becsléssel.

A következő témánk az, hogy hogyan lehet jó paraméterbecsést adni. A maximum likelihood becslést fogom röviden ismertetni. Ennek két jó tulajdonsága is van. Egyrészt egy általános elvet fogalmaz meg, aminek a segítségével bonyolult esetekben is tudunk javasolni egy módszert az ismeretlen paraméter(ek) becslésére. Másrészt ez egy aszimptotikusan optimális becslést biztosít bizonyos regularitási feltételek teljesülése esetén. Olyan kis hibával becsüli meg a keresett paramétert nagy minta elemszám esetén, mint amilyet a Kramer–Rao egyenlőtlenség sugall.

Heurisztikus szinten ismertetem azon eredmény bizonyítását, amely megadja a maximum likelihood becslés hibájának az aszimptotikus viselkedését nagy elemszámú minta esetén. Megadom a maximum likelihood egyenlet megoldásának aszimptotikus viselkedését bizonyos lineáris közelítéseket végrehajtva a valódi paraméter kis környezetében. Ezen közelítések elhanyagolhatóan kis hibát adnak alkalmas feltételek teljesülése esetén. Ennek a heurisztikus szinten plauzibilis állításnak a bizonyítása adja a precíz bizonyítást, amit elhagyok. Ehelyett megelégszem egy a Bolla–Krámli könyvben (szintén bizonyítás nélkül) kimondott eredmény ismertetésével.

Először csak a legegyszerűbb esetet vizsgálom, amikor egy egydimenziós paraméter becslését tekintem. A következő problémával foglalkozom. Legyen adva $f(x,\vartheta)$ sűrűségfüggvények egy családja a számegyenesen, amelynek elemei egy $\vartheta\in R^1$, azaz értékét a számegyenesen felvevő paramétertől függnek, és figyeljük meg $\xi_1(\omega),\dots,\xi_n(\omega)$ független, egyforma eloszlású valószínűségi változók egy sorozatát, amelyek eloszlásának van egy $f(x,\vartheta_0)$, sűrűségfüggvénye, ahol ϑ_0, ismeretlen paraméter. Ezt a paramétert akarjuk megbecsülni a megfigyelt valószínűségi változók segítségével.

A maximum likelihood módszer a következő becslési eljárást javasolja. Legyen az a $\hat\vartheta_n=\hat\vartheta_n(\xi_1,\dots,\xi_n)$ szám a ϑ_0 paraméter maximum likelihood becslése, amelyikre igaz az, hogy a $(\xi_1,\dots,\xi_n)$ vektor sűrűségfüggvénye, azaz a

$$\prod_{k=1}^n f(\xi_k,\vartheta)=\exp\left\{\sum_{k=1}^n \log f(\xi_k,\vartheta)\right\}$$

szorzat itt veszi fel a maximumát. Ez a szám megtalálható, mint az

$$\sum_{k=1}^n \frac{\partial}{\partial\vartheta}\log f(\xi_k,\vartheta)=0 \tag{9}$$

úgynevezett maximum likelihood egyenlet megoldása. Ezután minket a $\hat\vartheta_n-$

ϑ_0 különbség aszimptotikus viselkedése érdekel, ahol $\hat{\vartheta}_n$ a (9) egyenlet (megfelelő) megoldása.

Ezen egyenlet közvetlen vizsgálata meglehetősen nehéz. de a (9) egyenlet baloldalán szereplő kifejezés baloldalának a Taylor sorfejtése az (ismeretlen) ϑ_0 pont körül jó és egyszerű approximációt ad a $\hat{\vartheta}_n$ mennyiségre, és lehetővé teszi a $\hat{\vartheta}_n - \vartheta_0$ különbség aszimptotikus viselkedésének a megadását.

Ez a Taylor sorfejtés azt adja, hogy

$$
\begin{aligned}
&\sum_{k=1}^{n} \frac{\partial}{\partial\vartheta} \log f(\xi_k, \hat{\vartheta}_n) = \sum_{k=1}^{n} \frac{\frac{\partial}{\partial\vartheta} f(\xi_k, \vartheta_0)}{f(\xi_k, \vartheta_0)} \\
&+(\hat{\vartheta}_n - \vartheta_0) \left(\sum_{k=1}^{n} \left(\frac{\frac{\partial^2}{\partial\vartheta^2} f(\xi_k, \vartheta_0)}{f(\xi_k, \vartheta_0)} - \frac{\left(\frac{\partial}{\partial\vartheta} f(\xi_k, \vartheta_0)\right)^2}{f^2(\xi_k, \bar{\vartheta}_0)} \right) \right) + O\left(n(\hat{\vartheta}_n - \vartheta_0)^2\right) \\
&= \sum_{k=1}^{n} \left(\eta_k + \zeta_k(\hat{\vartheta}_n - \vartheta_0)\right) + O\left(n(\hat{\vartheta}_n - \vartheta_0)^2\right),
\end{aligned} \tag{10}
$$

ahol

$$
\eta_k = \frac{\frac{\partial}{\partial\vartheta} f(\xi_k, \vartheta_0)}{f(\xi_k, \vartheta_0)}, \quad \text{és} \quad \zeta_k = \frac{\frac{\partial^2}{\partial\vartheta^2} f(\xi_k, \vartheta_0)}{f(\xi_k, \vartheta_0)} - \frac{\left(\frac{\partial}{\partial\vartheta} f(\xi_k, \vartheta_0)\right)^2}{f^2(\xi_k, \bar{\vartheta}_0)}
$$

a $k = 1, \dots, n$ indexekre. Meg akarjuk érteni a (10) kifejezés jobboldalán szereplő mennyiség aszimptotikus viselkedését.

Az

$$
E\eta_k = \int \frac{\frac{\partial}{\partial\vartheta} f(x, \vartheta_0)}{f(x, \vartheta_0)} f(x, \vartheta_0)\, dx = \frac{\partial}{\partial\vartheta} \int f(x, \vartheta_0)\, dx = 0
$$

azonosság érvényes, mert $\int f(x, \vartheta)\, dx = 1$ minden ϑ paraméterre, és ezen reláció differenciálása adja az utolsó azonosságot. Hasonlóan $E\eta_k^2 = -E\zeta_k = \int \frac{\left(\frac{\partial}{\partial\vartheta} f(x,\vartheta_0)\right)^2}{f(x,\vartheta_0)}\, dx > 0$, $k = 1, \dots, n$. Ezért a centrális határeloszlástétel szerint $\chi_n = \frac{1}{\sqrt{n}} \sum\limits_{k=1}^{n} \eta_k$ aszimptotikusan normális nulla várható értékkel és $I(\vartheta) = \int \frac{\left(\frac{\partial}{\partial\vartheta} f(x,\vartheta_0)\right)^2}{f(x,\vartheta_0)}\, dx > 0$ szórásnégyzettel. A matematikai statisztika irodalomban az $I(\vartheta)$ számot Fisher információnak hívják. A nagy számok törvénye szerint $\frac{1}{n} \sum\limits_{k=1}^{n} \zeta_k \sim -I(\vartheta)$.

Így a (10) formula a $\hat{\vartheta}_n$ maximum likelihood becslés approximációját sugallja azzal a $\tilde{\vartheta}_n$ valószínűségi változóval, amelyet az $\tilde{\vartheta}_n - \vartheta_0 = -\frac{\sum\limits_{k=1}^n \eta_k}{\sum\limits_{k=1}^n \zeta_k}$ egyenlet határoz meg, és az előző számolásokból következik, hogy $\sqrt{n}(\tilde{\vartheta}_n - \vartheta_0)$ aszimptotikusan normális nulla várható értékkel és $\frac{1}{I(\vartheta)}$ szórásnégyzettel. A $\tilde{\vartheta}_n$ valószínűségi változó nem megoldása a (9) egyenletnek, az ezen egyenlet baloldalán szereplő kifejezés értékének a nagyságrendje $O(n(\tilde{\vartheta}_n - \vartheta_0)^2) = O(1)$ ebben a pontban. Másrészt, némi számolás mutatja, hogy a (9) egyenlet baloldalán szereplő függvény differenciálhányadosa ebben a pontban viszonylag nagy, nagyobb, mint const. n valamilyen const. > 0 számmal. Innen következik, hogy a maximum likelihood egyenletnek van egy olyan $\hat{\vartheta}_n$ megoldása, amelyre $\hat{\vartheta}_n - \tilde{\vartheta}_n = O\left(\frac{1}{n}\right)$. Ezért a $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ és $\sqrt{n}(\tilde{\vartheta}_n - \vartheta_0)$ kifejezéseknek hasonló az aszimptotikus viselkedésük.

A fent ismertetett bizonyításvázlat lényege igy foglalható össze: Vegyük egy egyszerű linearizált verzióját a vizsgálandó kifejezésnek egy alkalmas Taylor sorfejtés segítségével, adjuk meg e linearizált verzió aszimptotikus viselkedését, és mutassuk meg, hogy a linearizáció csak elhanyagolhatóan kis hibát okozott.

Ismertetek egy pontos feltételekkel megfogalmazott eredményt a maximum likelihood becslés viselkedéséről. Ez Cramer és Dugué tétele, amely a Bolla–Krámli könyv 4.3 tételében van kimondva.

Cramer–Dugué tétel. *Legyen adva egy $X = (X_1, \dots, X_n)$ véletlen vektor egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, ahol $X_1, \dots . X_n$ független, egyforma eloszlású valószínűségi változók, amelyek eloszlása P_ϑ valamilyen $f_\vartheta(x)$, $\vartheta \in \Theta$, sűrűségfüggvénnyel, ahol a Θ paraméterhalmaz a számegyenes nyílt részhalmaza, és $f_\vartheta \neq f_{\vartheta'}$, ha $\vartheta \neq \vartheta'$. Tegyük fel, hogy teljesülnek a következő feltételek.*

i.) A ϑ_0 valódi paraméter egy U környezetében léteznek a $f_\vartheta(x)$ sűrűségfüggvénynek a ϑ paraméter szerinti 1. 2. és 3. parciális deriváltjai.

ii.) Léteznek olyan F_1, F_2, F_3 függvények, amelyekre

$$\left|\frac{\partial}{\partial\vartheta} f_\vartheta(x)\right| < F_1(x), \qquad \left|\frac{\partial^2}{\partial\vartheta^2} f_\vartheta(x)\right| < F_2(x), \qquad \left|\frac{\partial^3}{\partial\vartheta^3} \log f_\vartheta(x)\right| < F_3(x)$$

minden $\vartheta \in U$ paraméterre. Továbbá F_1 és F_2 integrálható függvények, és $E_{\vartheta_0}(F_3(x)) < \infty$.

iii.) $I(\vartheta) < \infty$ minden $\vartheta \in U$ paraméterre, és $I(\vartheta_0) > 0$.

Akkor az n elemű mintához tartozó maximum likelihood egyenletnek létezik olyan $\hat{\vartheta}_n$ gyöke, amelyre $n \to \infty$ esetén

(1) $P(\hat{\vartheta}_n \to \vartheta_0) = 1$.

(2) $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \to \mathcal{N}\left(0, \frac{1}{I(\vartheta_0)}\right)$ *eloszlásban, ahol* $\mathcal{N}\left(0, \frac{1}{I(\vartheta_0)}\right)$ *a normális eloszlást jelöli* 0 *várható értékkel és* $\frac{1}{I(\vartheta_0)}$ *szórásnégyzettel.*

(3) $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ *minden momentuma konvergál az* $\mathcal{N}\left(0, \frac{1}{I(\vartheta_0)}\right)$ *eloszlás által meghatározott megfelelő momentumhoz, ha $n \to \infty$.*

Végül röviden ismertetem a maximum likelihood becslés aszimptotikus viselkedéséről szóló eredményt abban az esetben, ha a Θ paramétertartomány a k-dimenziós tér nyílt részhalmaza, $k \geq 1$. Az eredménynek most is csak egy nagyon vázlatos heurisztikus magyarázatát fogom adni.

Annak érdekében, hogy kimondjuk az eredményt szükséges az eddigiektől némileg eltérő jelölést bevezetni. Jelölje $\vartheta = (\vartheta_1, \dots, \vartheta_k)$ a Θ paramétertartomány egy pontját, $\vartheta(0) = (\vartheta_1(0), \dots, \vartheta_k(0))$ az (ismeretlen) valódi becsülendő paramétert, és legyen $\hat{\vartheta}(n) = (\hat{\vartheta}_1(n), \dots, \hat{\vartheta}_k(n))$ az n elemű minta segítségével elvégzett maximum likelihood becslés értéke.

Jelölje $\xi_1, \dots, \xi_n$ a megfigyelt n elemű mintát. Ekkor a k-dimenziós paramétertartomány esetében a maximum likelihood egyenlet a következő alakú.

$$\sum_{j=1}^{n} \frac{\partial_l}{\partial \vartheta_l} \log f(\xi_j, \vartheta) = 0, \quad l = 1, \dots, k.$$

A $\vartheta(0)$ pont körüli Taylor sorfejtéssel kapjuk, hogy a fenti összeg értéke a $\hat{\vartheta}(n)$ pontban

$$\sum_{j=1}^{n} \frac{\partial}{\partial \vartheta_l} \log f(\xi_j, \hat{\vartheta}(n)) = \sum_{j=1}^{n} \frac{\frac{\partial_l}{\partial \vartheta_l} f(\xi_j, \vartheta(0))}{f(\xi_j, \vartheta(0))}$$
$$+ \sum_{j=1}^{n} \sum_{m=1}^{k} (\hat{\vartheta}_m(n) - \vartheta_m(0)) \left(\frac{\frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} f(\xi_j, \vartheta(0))}{f(\xi_j, \vartheta(0))} \right.$$

$$
\begin{aligned}
&\qquad\qquad\qquad - \frac{\frac{\partial}{\partial\vartheta_l} f(\xi_j,\vartheta(0)) \frac{\partial}{\partial\vartheta_m} f(\xi_j,\vartheta(0))}{f^2(\xi_j,\vartheta(0))}\Bigg) \\
&\quad +O\left(n(\hat\vartheta(n)-\vartheta(0))^2\right) \\
&= \sum_{j=1}^n \left(\eta_{j,l}(n) + \left(\sum_{m=1}^k \zeta_{j,l,m}(n)(\hat\vartheta_m(n)-\vartheta_m(0))\right)\right) \\
&\quad +O\left(n(\hat\vartheta(n)-\vartheta(0))^2\right),
\end{aligned}
$$

$l=1\dots,k$, ahol

$$
\eta_{j,l}(n) = \frac{\frac{\partial}{\partial\vartheta_l} f(\xi_j,\vartheta(0))}{f(\xi_j,\vartheta(0))},
$$

és

$$
\zeta_{j,l,m}(n) = \frac{\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m} f(\xi_j,\vartheta(0))}{f(\xi_j,\vartheta(0))} - \frac{\frac{\partial}{\partial\vartheta_l} f(\xi_j,\vartheta(0)) \frac{\partial}{\partial\vartheta_m} f(\xi_j,\vartheta(0))}{f^2(\xi_j,\vartheta(0))}
$$

a $j=1,\dots,n$ és $l,m=1,\dots,k$ indexekre.

Innen az elhanyagolható hibatagokat elhagyva a következő jó közelítést kapjuk a $\hat\vartheta_m(n)-\vartheta_m(0)$, $m=1,\dots,k$, kifejezésekre a maximum likelihood becslés, azaz az n elemű minta segítségével adott $\bar\vartheta(n)$ becslés és a valódi $\vartheta(0)$ paraméter különbségére. Vezessük be az $U_l(n)=\sum_{j=1}^n \eta_{j,l}(n)$ és $V_{l,m}(n)=\sum_{j=1}^n \zeta_{j,l,m}(n)$ mennyiségeket. Segítségükkel felírhatjuk az

$$
\sum_{m=1}^k V_{l,m}(n)(\hat\vartheta_m(n)-\vartheta_m(0)) = -U_l(n), \quad l=1,\dots,k,
$$

lineáris egyenletrendszert. Innen

$$
(\hat\vartheta_1(n)-\vartheta_1(0),\dots,(\hat\vartheta_k(n)-\vartheta_k(0)) = -(U_1(n),\dots,U_k(n))Z(n)^{-1}, \tag{11}
$$

ahol $Z(n)$ a $Z(n)=(V_{l,m}(n))$, $l,m=1,\dots,k$ mátrix.

Az egydimenziós esethez hasonló számolás adja, hogy $EU_l=0$, $l=1,\dots,k$, és az $\frac{1}{\sqrt n}(U_1(n),\dots,U_k(n))$ vektor a centrális határeloszlástétel szerint eloszlásban konvergál egy nulla várható értékű és $I(\vartheta_0)$ kovarianciamátrixú normális eloszláshoz, ahol $I(\vartheta_0)$ a Fisher információ mátrix. Másrészt a nagy számok törvénye alkalmazható mindegyik $V_{l,m}(n)$, $l,m=1,\dots,k$, valószínűségi változóra, és az azt adja, hogy az $\frac1n(V_{l,m}(n))=\frac1n Z(n)$, $l,m=1,\dots,k$, mátrix egy valószínűséggel konvergál az $I(\vartheta_0)$ mátrixhoz, ezért inverze konvergál e mátrix inverzéhez.

Ezen eredmények alapján nem nehéz belátni, felhasználva a (11) formulát, hogy a $\sqrt{n}(\bar{\vartheta}(n) - \vartheta(0))$ (normalizált) maximum likelihood becslés hibája aszimptotikusan nulla várható értékű normális eloszlású valószínűségi változó, amelynek kovarianciája

$$\begin{aligned}
&\lim_{n\to\infty} nE[(U_1(n),\dots,U_k(n))Z(n)^{-1}]^*[(U_1(n),\dots,U_k(n))Z(n)^{-1}]\\
&\qquad = I(\vartheta_0)^{-1}E\left(\frac{1}{n}[(U_1(n),\dots,U_k(n))]^*(U_1(n),\dots,U_k(n))\right)I(\vartheta_0)^{-1}\\
&\qquad = I(\vartheta_0)^{-1},
\end{aligned}$$

mert $\lim\limits_{n\to\infty} nZ(n)^{-1} = I(\vartheta_0)^{-1}$, és

$$\lim_{n\to\infty} E\left(\frac{1}{n}[(U_1(n),\dots,U_k(n))]^*(U_1(n),\dots,U_k(n))\right) = I(\vartheta_0).$$

Ez azt jelenti, hogy szép, reguláris esetekben a normalizált maximum likelihood becslés hibája, azaz a $\sqrt{n}(\bar{\vartheta}(n) - \vartheta(0))$ kifejezés aszimptotikusan normális eloszlású, nulla várható értékkel és $I(\vartheta_0)^{-1}$ kovariancia mátrixszal.

Végül összehasonlítom a maximum likelihood becslés hibájáról kapott eredményt a Kramer–Rao egyenlőtlenség állításával abban az esetben, ha k-dimenziós reguláris modelleket tekintünk.

Legyen adva független, egyforma eloszlású $\xi_1,\dots,\xi_n$ valószínűségi változók sorozata $f_{\vartheta(0)}(x)$ sűrűségfüggvénnyel, ahol $\vartheta(0)\in\Theta$, létezik $f_\vartheta(x)$ sűrűségfüggvény minden $\vartheta\in\Theta$ paraméterre, azok szépen viselkednek, és Θ egy nyílt halmaz a k-dimenziós Euklideszi térben. Tekintsük a $\vartheta(0)$ paraméter $\hat{\vartheta}(n)$ maximum likelihood becslését. Legyen $\psi(\cdot)$ egy sima függvény a Θ halmazon, és vizsgáljuk a $\psi(\hat{\vartheta}(n)) - \psi(\vartheta(0))$ különbség aszimptotikus viselkedését nagy n paraméterre.

Mivel $\hat{\vartheta}(n)) - \vartheta(0))$ kicsi, (nagyságrendje $O(n^{-1/2})$), ezért a

$$Z(n) = \sum_{l=1}^{k} \frac{\partial}{\partial\vartheta_l}\psi(\vartheta(0))((\hat{\vartheta}_l(n)) - \vartheta_l(0))$$

kifejezés jó közelítése a $\psi(\hat{\vartheta}(n)) - \psi(\vartheta(0))$ különbségnek. A $\sqrt{n}Z(n)$ és $\sqrt{n}(\psi(\hat{\vartheta}(n)) - \psi(\vartheta(0)))$ kifejezéseknek ugyanaz a határeloszlása. Viszont $\sqrt{n}Z(n)$ határeloszlását tudjuk a maximum likelihood becslésre vonatkozó centrális határeloszlástétel alapján. Valóban, abból, hogy $\sqrt{n}(\hat{\vartheta}(n) - \vartheta(0))$

határeloszlása a k-dimenziós normális eloszlás $I(\vartheta(0))^{-1}$ kovariancia mátrixszal és nulla várható értékkel, következik, hogy $\sqrt{n}Z(n)$ határeloszlása a normális eloszlás nulla várható értékkel és $\sigma^2(\vartheta(0)$ szórásnégyzettel, ahol

$$\begin{aligned}\sigma^2(\vartheta(0)) \ &= \left(\frac{\partial}{\partial\vartheta_1}\psi(\vartheta(0)),\dots,\frac{\partial}{\partial\vartheta_k}\psi(\vartheta(0))\right) I(\vartheta(0))^{-1} \\ &\qquad\qquad\qquad\qquad \left(\frac{\partial}{\partial\vartheta_1}\psi(\vartheta(0)),\dots,\frac{\partial}{\partial\vartheta_k}\psi(\vartheta(0))\right)^* .\end{aligned}$$

Ezért $\sqrt{n}(\psi(\hat\vartheta(n)) - \psi(\vartheta(0)))$ határeloszlása is a normális eloszlás nulla várható értékkel és $\sigma^2(\vartheta(0))$ szórásnégyzettel.

Érdemes ezt az eredményt összehasonlítani a Cramer–Rao egyenlőtlenséggel. Utóbbi azt állítja hogy a $\psi(\vartheta(0))$ függvény bármely torzítatlan $\bar\vartheta(n) = T(\xi_1,\dots,\xi_n)$ becslésére $\mathrm{Var}\left(\sqrt{n}[\bar\vartheta(n) - \psi(\vartheta(0))]\right) \ge \sigma^2(\vartheta(0))$. Ez azt jelenti, hogy a $\psi(\hat\vartheta(n))$ maximum likelihood becslés szórásnégyzete aszimptotikusan nem nagyobb, mint $\psi(\vartheta(0))$ bármely torzítatlan becsléséé.

Kiegészítés.

A most tárgyalt maximum likelihood becslés elve általánosabb feltételek mellett is alkalmazható. A módszer lényege az, hogy olyan paramétert keresünk, amelyre a hozzá tartozó eloszlás esetén a megfigyelt minta sűrűségfüggvényének az értéke az adott helyen a lehető legnagyobb. Lehet, hogy nem abszolút, hanem csak lokális maximumot keresünk. Az ismertetett vázlatos bizonyítással indokolt eredmény arról szólt, hogy szép esetekben ezt a szélsőértéket, tehát a maximum likelihood becslést a sűrűségfüggvény értékének a paraméter szerinti deriválása segítségével megkapjuk, és ilyen módon egy nagyon jó tulajdonságokkal rendelkező becsléshez jutunk. De ezt a módszert olyan esetekben is alkalmazhatjuk, amikor az ebben a tárgyalásban felhasznált regularítási feltételek nem teljesülnek, például a maximumot nem tudjuk differenciálás segítségével megtalálni. De a maximum likelihood becslés elve ilyenkor is alkalmazható, tehát a keresett maximumot ilyenkor is megkereshetjük, és a kapott maximumot tekinthetjük a maximum likelihood becslés értékének. Ez az eljárás gyakran ad jó eredményt. Az alábbi példa tanulságos lehet.

Legyen adva egy n elemű független egyforma eloszlású valószínűségi változókból álló $X = (X_1,\dots,X_n)$ minta, amelynek X_j, $1 \le j \le n$, elemei egyenletes eloszlásúak egy $[\vartheta,\vartheta+1]$ intervallumban, ahol $\vartheta\in R^1$ ismeretlen paraméter, és ezt akarjuk megbecsülni megfigyelésünk alapján. Értsük meg, hogy milyen módszert javasol a maximum likelihood becslés elve ebben az esetben.

Ekkor a megfigyelt minta eloszlásának van sűrűségfüggvénye (a Lebesgue mérték szerint), amely így írható fel: $f_\vartheta(x_1,\dots,x_n)=1$, ha $\vartheta\le x_j\le\vartheta+1$ minden $1\le j\le n$ indexre, és $f_\vartheta(x_1,\dots,x_n)=0$ egyébként. Érdemes ezt olyan egyszerűbb alakban felírni, amelyben ez az $f_\vartheta(\cdot)$ sűrűségfüggvény nem n hanem csak két változó függvénye. Ennek érdekében vezessük be a a következő T: $R^n\to R^2$ leképezést.

$$T(x_1,x_2,\dots,x_n)=(\min(x_1,\dots,x_n),\max(x_1,\dots,x_n)).$$

Definiáljuk továbbá a következő g_ϑ: $R^2\to R^1$ függvényt. $g_\vartheta(x,y)=1$, ha $\vartheta\le x\le y\le\vartheta+1$, és $g_\vartheta(x,y)=0$ egyébként.

Ezzel a választással érvényes az

$$f_\vartheta(x_1,\dots,x_n)=g_\vartheta(T(x_1,\dots,x_n))$$

azonosság.

Az $f_\vartheta(x_1,\dots,x_n)$ függvénynek a maximumhelyét úgy keressük, mint a $g_\vartheta(T(x_1,\dots,x_n))$ függvény maximumhelyét. Ez a maximumhely nem egyértelmű. Olyan $\hat\vartheta$ értékekre vétetik fel, amelyekre $y-1\le\hat\vartheta\le x$, ahol

$$x=\min(x_1,\dots,x_n),\quad y=\max(x_1,\dots,x_n).$$

Ilyen $\hat\vartheta$ értékekre lesz $g_{\hat\vartheta}(T(x_1,\dots,x_n))=1$. Ez azt jelenti, hogy a maximum likelihood becslés ebben az esetben nem egyértelmű. Definiáljuk a legkisebb $X_{(1)}=\min(X_1,\dots,X_n)$ és legnagyobb $X_{(n)}=\max(X_1,\dots,X_n)$ elemet az $X=(X_1,\dots,X_n)$ mintából. Egy maximum likelihood becslést úgy kapunk, mint ezen $X_{(1)}$ és $X_{(n)}$ elemek $S(X_{(1)},X_{(n)})$ függvényét alkalmas $S(x,y)$ függvénnyel. Az $S(X_{(1)},X_{(n)})$ valószínűségi változó akkor lesz maximum likelihood becslés, ha $y-1\le S(x,y)\le x$ az $S(x,y)$ függvényre minden olyan (x,y) pontpárra, amelyre $x\le y\le x+1$.

Ezen maximum likelihood becslések közül a legtermészetesebb a $\hat\vartheta_n=\frac{X_{(1)}+X_{(n)}}{2}-\frac12$ becslés, azaz a $\hat\vartheta_n=S(X_{(1)},X_{(n)})$ becslés az $S(x,y)=\frac{x+y}{2}-\frac12$ függvénnyel. Ez analóg a várható értéknek a mintaelemek átlaga szerinti becslésével, ami jó becslés reguláris esetekben, de itt másfajta átlagot érdemes venni. Vegyük észre, hogy a Neyman–Fisher faktorizációs tétel alapján $T(X_{(1)}, X_{(n)})$ elégséges statisztika. (Általában a maximum likelihood becslések elégséges statisztikán alapulnak, és ezt nem nehéz látni a Neyman–Fisher faktorizációs tétel segítségével.) Lássuk be, hogy ez a becslés torzítatlan, és számoljuk ki a szórásnégyzetét.

Feladatok:

Bizonyítsa be, hogy ha $X_1, \dots, X_n$ független, egyforma eloszlású valószínűségi változók egyenletes eloszlással a $[\vartheta, \vartheta+1]$ intervallumban, akkor az $(X_{(1)}, X_{(n)})$ véletlen vektor sűrűségfüggvénye, ahol $X_{(1)} = \min(X_1, \dots, X_n)$ és $X_{(n)} = \max(X_1, \dots, X_n)$ az a $g(x,y)$ függvény, amelyre

$$g(x,y) = n(n-1)(y-x)^{n-2}, \text{ ha } \vartheta \le x \le y \le \vartheta+1, \text{ és } g(x,y) = 0 \text{ egyébként.}$$

Mutassa meg, hogy a $\hat{\vartheta}_n = \frac{X_{(1)}+X_{(n)}}{2} - \frac{1}{2}$ becslés torzítatlan, és szórásnégyzete $\frac{1}{2(n+1)(n+2)}$.

Megjegyzés. S Cramer–Rao egyenlőtlenség alapján bizonyos regularitási feltételek teljesülése esetében egy torzítatlan becslés hibájának a szórásnégyzete nagyobb, mint $\frac{\text{const.}}{n}$. Ebben a példában ez a szórásnégyzet $O(\frac{1}{n^2})$, viszont nem teljesülnek a Cramer–Rao egyenlőtlenség feltételei.

Segítség. A következő integrált kell kiszámolni a $\hat{\vartheta}_n = \frac{X_{(1)}+X_{(n)}}{2} - \frac{1}{2}$ becslés szórásnégyzetének kiszámításához:

$$\int_{\{(x,y)\colon\, \vartheta \le x \le y \le \vartheta+1\}} n(n-1)(y-x)^{n-2} \left(\frac{(x+y)}{2} - \frac{1}{2} - \vartheta \right)^2 dx\, dy.$$

Érdemes az $u = y - x$, $v = x + y - 1 - 2\vartheta$ változó transzformációt elvégezni, amelynek Jacobi determinánsa 2. Az előző integrál a következőképp fejezhető ki az u és v változók segítségével:

$$\frac{n(n-1)}{2} \int_{\{(u,v) \in D\}} u^{n-2} \cdot \frac{v^2}{4}\, du\, dv,$$

ahol $D = \{(u,v)\colon\ 0 \le u \le 1,\ u-1 \le v \le 1-u\}$.

Tekintsük a következő feladatot. Legyenek $X_1, \dots, X_n$ független, egyforma eloszlású valószínűségi változók egyenletes eloszlással a $[0, \vartheta]$ intervallumban valamilyen $\vartheta > 0$ paraméterrel. Határozza meg a ϑ paraméter maximum likelihood becslését az $X = (X_1, \dots, X_n)$ minta alapján, és számolja ki e becslés hibájának várható értékét és szórásnégyzetét.

Segítség. Először a maximum likelihood becslés meghatározásával foglalkozunk. Ennek módszerét ismertetem röviden. Az $X = (X_1, \dots, X_n)$ minta sűrűségfüggvénye $f_\vartheta(x_1, \dots, x_n) = \vartheta^{-n}$, ha $0 \le x_j \le \vartheta$, $1 \le j \le n$, és $f_\vartheta(x_1, \dots, x_n) = 0$ egyébként. Írjuk ezt fel egyszerűbb, számunkra hasznosabb alakban. Definiáljuk az $y = T(x_1, \dots, x_n)$ függvényt a következő módon. Legyen

$$T(x_1, \dots, x_n) = \max(x_1, \dots, x_n), \text{ ha } x_j \ge 0 \text{ minden } 1 \le j \le n \text{ indexre},$$

$T(x_1, \dots, x_n) = 0$ egyébként. Legyen $g_\vartheta(x) = \vartheta^{-n}$, ha $0 < x \le \vartheta$, és $g_\vartheta(x) = 0$ egyébként, és legyen $X_{(n)} = \max(X_1, \dots, X_n)$. Ekkor

$$f_\vartheta(x_1, \dots, x_n) = g_\vartheta(T(x_1, \dots, x_n)),$$

és ez maximumát a $\vartheta = T(x_1, \dots, x_n)$ paraméterre veszi fel. Ezért a maximum likelihood becslés a $\hat\vartheta = T(X_1, \dots, X_n) = X_{(n)}$ becslés. Jegyezzük meg, hogy $X_{(n)}$ sűrűségfüggvényét a $h_\vartheta(x) = nx^{n-1}\vartheta^{-n}$, ha $0 \le x \le \vartheta$, és $h_\vartheta(x) = 0$ egyébként képlet határozza meg. Ez lehetővé teszi a a $\hat\vartheta_n = X_{(n)}$ maximum likelihood becslés várható értékének és szórásnégyzetének kiszámítását.

Tekintsük a következő feladatot. Határozzuk meg egy normális eloszlású, azaz $f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ sűrűségfüggvényű valószínűségi változó μ és σ paraméterét egy $X = (X_1, \dots, X_n)$ minta ismeretében a maximum likelihood becslés módszerével. Tekintsük azt az esetet, amikor mind a μ mind a σ paraméter ismeretlen, és a (μ, σ) párt akarjuk megbecsülni, és azt az esetet is, amikor az egyik paraméter értéke ismert, és a másik paramétert akarjuk megbecsülni.