A december 5-i szeminárium témája
Rövid összefoglaló

A matematikai statisztika rendkívül fontos témái a becsléselmélet, a hipotézisvizsgálat és konfidenica intervallumok meghatározása. A gyakorlat utolsó két óráján ezzekkel a témákkal foglalkozunk. Ennek a gyakorlatnak a témája a becsléselmélet. A tipikus becsléselméleti problémák a következő jellegűek.

Adott független, egyforma eloszlású valószínűségi változók egy $\xi_1, \dots, \xi_n$ sorozata, melyek eloszlása valamely $F_\vartheta(\cdot)$ eloszlásfüggvény valamilyen ismeretlen ϑ paraméterrel. Becsüljük meg ezt a ϑ paramétert, azaz az $F_\vartheta(\cdot)$ eloszlásfüggvényt minél jobban. Formálisan, definiáljunk olyan $T_n(x_1, \dots, x_n)$ n-változós függvényt, melyre a megfigyelt véletlen $\xi_1, \dots, \xi_n$ sorozat segítségével kiszámítható $\hat{\vartheta}_n = T_n(\xi_1, \dots, \xi_n)$ valószínűségi változó valamilyen értelemben jól közelíti az ismeretlen ϑ paramétert.

Vizsgálhatjuk azt az általánosabb problémát, amikor az ismeretlen $F_{\vartheta_1,\dots,\vartheta_k}(\cdot)$ eloszlásfüggvény k ismeretlen $\vartheta_1, \dots, \vartheta_k$ paramétertől függ, (ahol k valamely rögzített egész szám,) és becsüljük meg az összes $\vartheta_1, \dots, \vartheta_k$ paramétert. Előfordulhat olyan eset is, amikor ezen ismeretlen paraméterek közül csak néhányat akarunk megbecsülni, de nem mindegyiket. Az ilyen jellegű problémákat nevezzük paraméteres becslési problémának. Előfordulhat, hogy a következő problémával foglalkozunk. Adott független, egyforma eloszlású valószínűségi változók egy $\xi_1, \dots, \xi_n$ sorozata, melyek eloszlása valamely ismeretlen $F(\cdot)$ eloszlásfüggvény. Jelen esetben nem tudunk semmit az ismeretlen eloszlásfüggvényről. Becsüljük meg a ξ_k valószínűségi változók várható értékét, szórásnégyzetét vagy az eloszlását. Az ilyen problémákat hívják nem paraméteres becslésproblémának.

Először tekintünk néhány példát.

1. Egy pénzdarabot sokszor egymás után feldobunk. Becsüljük meg annak a valószínűségét, hogy a pénzdarab a fej-oldalra esik.

   Írjunk 1-t fejdobás és 0-t írásdobás esetén. Ekkor a probléma formálisan a következő módon fogalmazható meg: Adott független $B(1,p)$ eloszlású (binomiális eloszlású, 1 és p paraméterekkel), valószínűségi változók $\xi_1, \dots, \xi_n$ sorozata (az egyes dobások eredményei), ahol a p paraméter ismeretlen. Becsüljük meg ezt az ismeretlen p paramétert.

   A természetes becslés, $\hat{p}_n = \frac{1}{n}\sum\limits_{k=1}^{n} \xi_k$, azaz $T_n(x_1, \dots, x_n) = \frac{1}{n}\sum\limits_{k=1}^{n} x_k$. Ekkor

   $$\hat{p}_n = T_n(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{k=1}^{n} \xi_k = \frac{\text{A fejdobások száma}}{n}.$$

   A nagy számok törvénye szerint $\hat{p}_n \Rightarrow p$, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl, azaz $\lim\limits_{n\to\infty} P(|\hat{p}_n - p| > \varepsilon) \to 0$ minden $\varepsilon > 0$ számra.

2. Arra vagyunk kiváncsiak, hogy egy dobókocka milyen valószínűséggel esik különböző oldalaira. Ennek meghatározása érdekében a dobókockát feldobjuk sokszor, és megjegyezzük, hogy a különböző dobások során melyik oldalára esett a kocka.

Formálisan megfogalmazva a probléma a következő: Adott $F_{p_1,\dots,p_6}$ eloszlássorozat, (ez az eloszláscsalád valójában csak 5 paramétertől függ, mert a $p_1 + \cdots + p_6 = 1$ reláció teljesül.) úgy, hogy egy $F_{p_1,\dots,p_6}$ eloszlású ξ valószínűségi változóra $P(\xi = j) = p_j$, $j = 1, 2, \dots, 6$. Becsüljük meg az ismeretlen p_j paramétereket. A természetes becslés a következő:

$$\hat{p}_j = \frac{\text{A } j \text{ eredményű dobások száma}}{n}, \quad 1 \le j \le 6.$$

Formálisan, $\hat{p}_j = T_j(\xi_1, \dots, \xi_n)$, $1 \le j \le n$, ahol

$$T_j(x_1, \dots, x_n) = \frac{\text{a } j \text{ értékek száma az } x_1, \dots, x_n \text{ sorozatban}}{n}.$$

A nagy számok törvénye alapján érvényes a $\hat{p}_j \Rightarrow p_j$, $1 \le j \le 6$ reláció, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl.

E feladat egy módosított változata: Az előző problémát tekintjük, de minket csak az érdekel, hogy milyen valószínűséggel esik a kocka a hatos oldalra. Ez egy természetes példa arra, amikor egy több paramétertől függő eloszláscsaládot figyelünk meg, de minket csak néhány, (jelen esetben egy) paraméter értéke érdekel.

3. Ledobunk egymástól függetlenül n pontot, melyek egy $[\vartheta, \vartheta+1]$ intervallumba esnek egyenletes eloszlással, azaz annak valószínűsége, hogy egy pont a $[\vartheta, \vartheta + 1]$ egy részintervallumába esik megegyezik ennek az intervallumnak a hosszával. Jelölje $\xi_1, \dots, \xi_n$ a ledobott pontok helyét. Becsüljük meg ezek ismeretében az ismeretlen ϑ paramétert.

   Erről a feladatról már volt szó a november 14-i gyakorlaton. Megbeszéltük, hogy két természetes becslés jelenik meg. Az egyik: $\tilde{\vartheta}_n = \dfrac{\xi_1 + \cdots + \xi_n}{n} - \dfrac{1}{2}$, a másik $\hat{\vartheta}_n = \dfrac{\xi_1^* + \xi_n^*}{n} - \dfrac{1}{2}$, ahol $\xi_1^* < \xi_2^* < \cdots < \xi_n^*$, a $\xi_1, \dots, \xi_n$ (véletlen) számok nagyság szerinti sorrendbe rakva. (Ezt hívják az irodalomban rendezett mintának.) Láttuk, hogy mind a két becslés jó, de a második becslés jobb az elsőnél a következő értelemben: $E\tilde{\vartheta}_n = \vartheta$, és $E\hat{\vartheta}_n = \vartheta$, azaz mind a két becslés várható értéke megegyezik a keresett paraméterrel. Viszont a várható érték körüli ingadozást természetes módon mérő szórásnégyzetekre a következő reláció érvényes: $\operatorname{Var} \tilde{\vartheta}_n = \dfrac{1}{12n}$, és $\operatorname{Var} \hat{\vartheta}_n = \dfrac{1}{2(n+1)(n+2)}$. Ezért $\operatorname{Var} \tilde{\vartheta}_n \ge \operatorname{Var} \hat{\vartheta}_n$ minden $n \ge 2$ számra.

4. Legyen egy lámpa élettartama exponenciális eloszlású valószínűségi változó ismeretlen λ paraméterrel, azaz $f(x) = \lambda e^{-\lambda x}$, ha $x \ge 0$, és $f(x) = 0$, ha $x \le 0$. Megvizsgáljuk n lámpa élettartamát. Ezek (egymástól független) $\xi_1, \dots, \xi_n$ exponenciális eloszlású valószínűségi változók λ paraméterrel. Becsüljük meg a λ paramétert.

   Első ránézésre nem tudunk kitalálni egy természetes becslést erre a λ paraméterre. Természetes kívánság egy olyan általános elv megfogalmazása, mely az általános esetben, például erre a problémára is egy jó módszert ad egy ismeretlen paraméter becslésére. Ilyen módszer az alább ismertetett maximum likelihood módszer. Ez általános elvet fogalmaz meg ismeretlen paraméter(ek) becslésére. Másik jó tulajdonsága a maximum likelihood módszernek az, hogy bizonyos itt nem ismertetett eredmények azt mondják ki, hogy a maximum likelihood módszer általános feltételek esetén nagy mintaelemszám esetén aszimptotikusan optimális eljárást nyújt.

Maximum likelihood módszer. *Legyen adva eloszlásfüggvényeknek egy λ vagy véges sok $(\lambda_1,\dots,\lambda_k)$ valós számal paraméterezett családja, melyeknek van*

$$f_\lambda(x)=f(\lambda,x) \quad \textit{vagy} \quad f_{\lambda_1,\dots,\lambda_k}(x)=f(\lambda_1,\dots,\lambda_k,x), \quad -\infty<x<\infty,$$

alakú sűrűségfüggvényük. Legyen $\xi_1,\dots,\xi_n$ független, egyforma eloszlású valószínűségi változók sorozata ezzel az $f_\lambda(x)=f(\lambda,x)$ vagy $f(\lambda_1,\dots,\lambda_k,x)$ sűrűségfüggvénnyel az ismeretlen λ paraméterrel vagy ismeretlen $(\lambda_1,\dots,\lambda_k)$ paramétercsaláddal. Ekkor a $\xi_1,\dots,\xi_n$ sorozat együttes sűrűségfüggvénye az $f(\lambda,x_1,\dots,x_n)=\prod\limits_{j=1}^n f(\lambda,x_j)$ illetve az $f(\lambda_1,\dots,\lambda_k,x_1,\dots,x_n)=\prod\limits_{j=1}^n f(\lambda_1,\dots,\lambda_k,x_j)$ n-dimenziós sűrűségfüggvény. Helyettesítsük be ebbe a paraméterektől függő sűrűségfüggvénybe a megfigyelt $\xi_1,\dots,\xi_n$ valószínűségi változókat, azaz tekintsük az

$$f(\lambda_1,\dots,\lambda_k,\xi_1,\dots,\xi_n) \quad \textit{illetve az} \quad f(\lambda_1,\dots,\lambda_k,\xi_1,\dots,\xi_n)$$

valószínűségi változót. Azt mondjuk, hogy a $\hat\lambda(\xi_1,\dots,\xi_n)$ (véletlen) szám maximum likelihood becslése a λ paraméternek, a $\hat\lambda_j(\xi_1,\dots,\xi_n)$, $1\le j\le k$, (véletlen) k dimenziós vektor maximum likelihood becslése a $(\lambda_1,\dots,\lambda_k)$ vektornak, ha ez a szám illetve vektor (lokális) maximuma az $f(\lambda,\xi_1,\dots,\xi_n)$ illetve az $f(\lambda_1,\dots,\lambda_k,\xi_1,\dots,\xi_n)$ függvénynek rögzített $\xi_1,\dots,\xi_n$ számra.

Megjegyzés: *A maximum likelihood becslésben lehet a sűrűségfüggvény maximuma helyett annak logaritmusát vizsgálni, azaz a $\log f(\lambda,\xi_1,\dots,\xi_n)=\sum\limits_{j=1}^n \log f(\lambda,\xi_j)$, illetve a $\log f(\lambda_1,\dots,\lambda_k,\xi_1,\dots,\xi_n)=\sum\limits_{j=1}^n \log f(\lambda_1,\dots,\lambda_k,\xi_j)$ függvény maximumát keresni. Ez a*

$$\sum_{j=1}^n \frac{d\log f(\lambda,\xi_j)}{d\lambda}=0$$

egyenlethez, illetve az

$$\sum_{j=1}^n \frac{\log f(\lambda_1,\dots,\lambda_k,\xi_j)}{\partial\lambda_j}=0, \quad 1\le j\le k,$$

egyenletrendszerhez vezet.

2. Megjegyzés: *A maximum likelihood becslés módszer természetes módon megfogalmazható abban az esetben is, amikor a vizsgált eloszlásfüggvénycsalád eloszlásfüggvényeinek nincs sűrűségfüggvényük, hanem például rácsos eloszlásúak. Ebben az esetben, ha az eloszlások bizonyos* m *értékeket vehetnek fel* $p_\lambda(m) = p(\lambda, m)$ *vagy* $p_{\lambda_1,\dots,\lambda_k}(m) = p(\lambda_1, \dots, \lambda_k, m)$ *valószínűséggel,* ($\sum\limits_m p(\lambda, m) = 1$, *illetve* $\sum\limits_m p(\lambda_1, \dots, \lambda_k, m) = 1$.) *Megfigyelünk egymástól független* $\xi_j = m_j$, $1 \le j \le n$, *értékeket, és a maximum likelihood elv alapján a* $\prod\limits_{j=1}^n p(\lambda, m_j)$, *illetve* $\prod\limits_{j=1}^n p(\lambda_1, \dots, \lambda_k, , m_j)$ *kifejezés maximumát adó* $\hat\lambda$ *illetve* $\hat\lambda_j$, $1 \le j \le k$, *értékek adják meg a maximum likelihood becslést.*

Számítsuk ki, hogy milyen becslést sugall a maximum likelihood módszer a most vizsgált probléma esetében.

Jelen esetben a $(\xi_1, \dots, \xi_n)$ vektor sűrűségfüggvénye a

$$\prod_{j=1}^n \lambda e^{-\lambda x_j} = \lambda^n \exp\left\{-\lambda \sum_{k=1}^n x_j\right\}$$

függvény. Ennek logaritmusa az $n \log \lambda - \lambda \sum\limits_{k=1}^n x_j$ függvény. Ezért a maximum likelihood módszer az $\dfrac{n}{\hat\lambda_n} = \sum\limits_{j=1}^n \xi_j$ egyenlethez és a $\hat\lambda_n = \dfrac{n}{\sum\limits_{j=1}^n \xi_j}$ becsléshez vezet. Jegyezzük meg, hogy egy λ paraméterű exponenciális eloszlású valószínűségi változó várható értéke $\dfrac{1}{\lambda}$. Ezért a nagy számok törvénye azt adja, hogy $\hat\lambda_n \Rightarrow \lambda$.

Nézzük meg milyen becslést sugall a maximum likelihood módszer a korábban tárgyalt problémákban. Az első feladatban, amikor egy pénzdarab fej oldalára való esésének a valószínűségét kell becsülnünk, jelöljük k-val a fejdobások (véletlen) számát n dobás esetén. Ekkor p paraméter esetén annak a valószínűsége, hogy pontosan k fejdobás következett be $P_p(k) = \binom{n}{k} p^k (1-p)^{n-k}$, ennek logaritmusa pedig $\log P_p(k) =$ const. $(k \log p + (n-k)\log(1-p))$, ahol a const. függhet a k számtól, de nem függ az ismeretlen p paramétertől. Ezért a maximum likelihood módszer által sugallt a p paraméter szerinti deriválás a $\dfrac{k}{\hat p_n} - \dfrac{n-k}{1-\hat p_n} = 0$ egyenlethez, és a $\hat p_n = \dfrac{k}{n}$ becsléshez vezet. Ez az eredetileg megadott becslés.

A második feladatban hasonló a helyzet. Annak a valószínűsége, hogy k_j j-es dobás következik be n kisérletben, $1 \le j \le 6$, $k_1 + \dots + k_6 = n$, p_j, $1 \le j \le 6$ esetében

$$P_{p_1,\dots,p_6} = \frac{n!}{k_1! \cdots k_6!} p_1^{k_1} \cdots p_6^{k_6} = \frac{n!}{k_1! \cdots k_6!} p_1^{k_1} \cdots p_5^{k_5} (1 - p_1 - \dots - p_5)^{k_6}.$$

E kifejezésben logaritmust véve, majd az így kapott kifejezésnek véve a parciális deriváltját a $p_1, \cdots, p_5$ változók szerint kapjuk a

$$\frac{k_j}{\hat{p}_{j,n}} = \frac{k_6}{1 - \hat{p}_{1,n} - \cdots - \hat{p}_{5,n}}, \quad 1 \le j \le 5,$$

egyenleteket. Ezeket az egyenleteket összehasonlítva (és felhasználva, hogy az egyenletek jobboldala nem függ a j indextől) kapjuk, hogy $\dfrac{\hat{p}_{j,n}}{\hat{p}_{s,n}} = \dfrac{k_j}{k_s}$ minden $1 \le j, s \le 6$ számra, azaz $C\hat{p}_{j,n} = k_j$, $1 \le j \le 5$, valamilyen ismeretlen C konstanssal. Ezt beírva az egyenletekbe kapjuk, hogy $k_6 = C - (k_1 + \cdots + k_5)$, ahonnan mivel $\sum\limits_{j=1}^{6} k_j = n$, $C = n$, és $\hat{p}_{j,n} = \dfrac{k_j}{n}$. Azaz a maximum likelihood módszer ebben az esetben is az eredetileg javasolt becslést adja. Megjegyezzük, hogy a fenti szélsőérték feladatot egyszerűbben kiszámíthattuk volna a feltételes szélsőérték kiszámítására kidolgozott Lagrange féle multiplikátor módszerrel.

A harmadik feladatban a $(\xi_1, \dots, \xi_n)$ véletlen vektor sűrűségfüggvénye 1, ha mindegyik $(\xi_1, \dots, \xi_n)$ pont a $[\vartheta, \vartheta + 1]$ intervallum belsejében van, azaz $\vartheta \le \xi_1^* < \xi_n^* \le \vartheta + 1$, egyébként pedig zéró. Ezért a vizsgálandó függvénynek nincs szigorú maximuma. Viszont szimmetriaokokból természetes azt a becslést tekinteni, melyre a $[\hat{\vartheta}_n, \hat{\vartheta}_n + 1]$ intervallum a $[\xi_1^*, \xi_n^*]$ intervallumot középen tartalmazza, és ez a tekintett $\hat{\vartheta}_n = \dfrac{\xi_1^* + \xi_n^*}{2} - \dfrac{1}{2}$ becslés.

5. Tekintsük a következő feladatot. Megfigyelünk $\xi_1, \dots, \xi_n$ független, azonos eloszlású valószínűségi változókat ismeretlen $F(x)$ eloszlásfüggvénnyel. Becsüljük meg:
   a. Az ismeretlen $F(x)$ eloszlásfüggvényt,
   b. A ξ_j valószínűségi változók várható értéket,
   c. A ξ_j valószínűségi szórásnégyzetét.

   Ezek a feladatok tipikus példái az úgynevezett nem parametrikus becsléselméletnek. Ugyanis nem lehet az összes lehetséges eloszlásfüggvényt természetes módon véges sok paraméterrel paraméterezni.

a. Az a.) esetben, az $F(x)$ eloszlásfüggvény becslése esetén természetes becslés az úgynevezett empirikus eloszlásfüggvény, azaz

$$F_n(x) = F_n(x, \xi_1, \dots, \xi_n) = \frac{\text{Azon } 1 \le k \le n \text{ indexek száma, melyekre } \xi_k < x}{n},$$
$$-\infty < x < \infty.$$

Be lehet látni, hogy $\sup\limits_{-\infty < x < \infty} |F_n(x) - F(x)| \Rightarrow 0$, ha $n \to \infty$ minden F eloszlásfüggvény esetén. Ez az órán is tárgyalt eredmény azt a tényt fejezi ki, hogy az empirikus eloszlásfüggvény valóban jó becslése az eloszlásfüggvénynek.

b. A várható érték természetes becslése $M_n = M_n(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{k=1}^{n} \xi_k$. Nem nehéz belátni, hogy $EM_n(\xi_1, \dots, \xi_n) = E\xi_1$. Továbbá, a nagy számok törvénye alapján $M_n(\xi_1, \dots, \xi_n) \Rightarrow E\xi_1$, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl.

c. A Szórásnégyzet szokásos becslése a következő becslés: $D_n^2 = D_n^2(\xi_1, \dots, \xi_n) = \frac{1}{n-1}\sum_{k=1}^{n} (\xi_k - M_n)^2$, ahol az $M_n = M_n(\xi_1, \dots, \xi_n)$ számot az előzőekben definiáltuk. Talán meglepő, hogy ebben a képletben $n-1$-gyel osztottunk és nem n-nel. Ennek egyébként nagy n mintaelemszám esetén kicsi a jelentősége. Ennek oka az, hogy mint azt némi (nem nehéz) számolás mutatja, ezzel a választással $ED_n^2(\xi_1, \dots, \xi_n) = \operatorname{Var}\xi_1$. Némi további számolás azt is mutatja, hogy $\lim\limits_{n\to\infty} \operatorname{Var} D_n^2(\xi_1, \dots, \xi_n) = 0$. Ezekből az eredményekből, illetve a Csebisev egyenlőtlenségből kapjuk, hogy $D_n^2(\xi_1, \dots, \xi_n) \Rightarrow \operatorname{Var}\xi_1$, ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl.

Végül megfogalmazunk néhány fontos fogalmat, elvet, melyek fontosak a becsléselméletben.

Legyen adva $F_\lambda(x)$, valószínűségi eloszlások egy családja, és adott n-re $\xi_1, \dots, \xi_n$ független, egyforma eloszlású valószínűségi változók sorozata, F_λ eloszlással, ahol λ ismeretlen paraméter. Azt mondjuk, hogy egy n-változós $T(x_1, \dots, x_n) = T_n(x_1, \dots, x_n)$ függvény becslést ad az ismeretlen λ paraméterre, ha az ismeretlen λ paramétert a $\hat{\lambda}_n = \hat{\lambda}_n(\xi_1, \dots, \xi_n) = T_n(\xi_1, \dots, \xi_n)$ valószínűségi változó segítségével becsüljük meg. A $\hat{\lambda}_n$, illetve e valószínűségi változó függvényeinek eloszlása függ az F_λ eloszlástól, ezért a λ paramétertől. Ezt jelölendő, a továbbiakban P_λ-val fogjuk jelölni azokat a valószínűségeket, melyek az $F_\lambda(\cdot)$ eloszlású $\xi_1, \dots, \xi_n$ valószínűségi változóktól függ.

Becslések konzisztenciája. *Az előbb bevezetett jelöléseket használva azt mondjuk, hogy a $T_n(x_1, \dots, x_n)$, $n = 1, 2, \dots,$ becslőfüggvények konzisztens becslést adnak, ha*

$$\hat{\lambda}_n = T_n(\xi_1, \dots, \xi_n) \Rightarrow \lambda \quad \textit{minden } \lambda \textit{ paraméterre,}$$

ahol $\Rightarrow$ sztochasztikus konvergenciát jelöl, azaz $\lim\limits_{n\to\infty} P_\lambda\left(|\hat{\lambda}_n - \lambda| > \varepsilon\right) = 0$ minden lehetséges λ paraméterre és $\varepsilon > 0$ számra.

Becslés torzítatlanságának definiciója. *Az előbb bevezetett jelöléseket használva azt mondjuk, hogy a $\hat{\lambda}_n = \hat{\lambda}_n(\xi_1, \dots, \xi_n) = T_n(x_1, \dots, x_n)$ becslés torzítatlan, ha*

$$E_\lambda \lambda_n(\xi_1, \dots, \xi_n) = \lambda \quad \textit{minden } \lambda \textit{ paraméterre.}$$

A $\hat{\lambda}_n = \hat{\lambda}_n(\xi_1, \dots, \xi_n) = T_n(x_1, \dots, x_n)$ becsléssorozat, $n = 1, 2, \dots,$ aszimptotikusan torzítatlan, ha

$$\lim_{n\to\infty} E_\lambda \lambda_n(\xi_1, \dots, \xi_n) = \lambda \quad \textit{minden } \lambda \textit{ paraméterre.}$$

A becsléselméletben gyakran keresnek olyan torzítatlan $\hat{\lambda}_n = \hat{\lambda}_n(\xi_1, \dots, \xi_n)$ becslést, melyre a $D_n(\lambda) = \operatorname{Var}_\lambda \hat{\lambda}_n(\xi_1, \dots, \xi_n)$ szórásnégyzet minél kisebb. Azt mondjuk, hogy egy torzítatlan $\hat{\lambda}_n$ becslés optimális, ha $\operatorname{Var}_\lambda \hat{\lambda}_n(\xi_1, \dots, \xi_n) \le \operatorname{Var}_\lambda \tilde{\lambda}_n(\xi_1, \dots, \xi_n)$ minden torzítatlan $\tilde{\lambda}_n(\xi_1, \dots, \xi_n)$ becslésre és λ paraméterre. Sok érdekes esetben van torzítatlan optimális becslés. Jegyezzük meg, hogy a Csebisev egyenlőtlenség alapján, ha egy $\hat{\lambda}_n = \hat{\lambda}_n(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots,$ becsléssorozat aszimptotikusan torzítatlan, és $\lim\limits_{n\to\infty} \operatorname{Var}_\lambda \hat{\lambda}_n(\xi_1, \dots, \xi_n) = 0$ minden λ paraméterre, akkor ez a becslés konzisztens. (Miért?) A maximum likelihood becslés nagyon általános feltételek mellett teljesíti ezeket a tulajdonságokat.