4. tétel

Hatásosabb becslés:

X $x_1, x_2, \ldots x_n$ $a(\underline{x}) = a_1 x_1 + \ldots + a_n x_n$

$a_1 + a_2 + \ldots + a_n = 1$ $a_i \geq 0$

$A = a_1 X + a_2 X_2 + \ldots + a_n X_n$

várható értéke

$E(A) = a_1 \cdot m + a_2 \cdot m + \ldots + a_n \cdot m = m \cdot (a_1 + a_2 + \ldots + a_n) = m$

amelyik becslésnek kisebb a varianciája, az jobban (hatásosabban) fog működni

$Var(A) = a_1^2 \sigma^2 + a_2^2 \sigma^2 + \ldots + a_n^2 \cdot \sigma^2 = \sigma^2 \cdot (a_1^2 + a_2^2 + \ldots + a_n^2)$ $D^2(x) = \sigma^2$

Cauchy-Schwarz egyenlőtlenséggel:

$|\underline{a} \cdot \underline{b}| \leq \sqrt{\underline{a}^2} \cdot \sqrt{\underline{b}^2}$ $(a_1 b_1 + \ldots + a_n b_n) \leq \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2} \cdot \sqrt{b_1^2 + \ldots + b_n^2}$

a, b vektor hossza

$\underline{b} = \underline{1}$ = összegző vektor $|a_1 + a_2 + \ldots + a_n| \leq \sqrt{a_1 + \ldots + a_n} \cdot \sqrt{n}$

$\frac{1}{n} \leq (a_1^2 + a_2^2 + \ldots + a_n^2)$

Minden $\frac{1}{n}$ lesz, ekkor $\frac{1}{n^2} + \frac{1}{n^2} + \ldots + \frac{1}{n^2} = \frac{1}{n}$ lineáris becslések körében a várható érték a legbiztosabb becslés

Fisher-információ és átparaméterezés. Tegyük fel, hogy a **T** paramétertér nyílt, összefüggõ részhalmaza a p dimenziós euklideszi térnek, és az X minta $f_t(x)$ sûrûségfüggvénye (t a paraméter) mm. x esetén differenciálható t-ben. Legyen **S** a q dimenziós tér nyílt részhalmaza és h : **S** $\rightarrow$ **T** differenciálható injektív leképezés (ekkor persze q $\leq$ p), jelölje h deriváltját H. Tekintsük a ***P*** ' = {$P_{h(s)}$: s eleme **S**-nek} részcsaládot, legyen P egy mérték ***P*** '-bõl: P = P_t = $P_{h(s)}$. Jelöljük a P-bõl vett minta Fisher-információját t-re ill. s-re I(t) ill. *I*(s). Ekkor

$$I(s) = H(s)^T I(h(s))\, H(s).$$

Megjegyzés: Ha ***P***-re teljesül (R), h folytonosan differenciálható és H teljes rangú minden t-re, akkor ***P*** '-re is teljesül (R).

Pontbecslések. Legyen g : **T** $\rightarrow$ $\mathbf{R}^k$, és a T k-dimenziós statisztika a g(t) egy becslése. Ekkor

- T *torzítása* : $b_T(t) = E_t(T) - g(t)$, T *torzítatlan*, ha a torzítása minden t-re 0, azaz a várható értéke éppen a becsülendõ mennyiség.
- Legyen W : $\mathbf{R}^k$ x **T** $\rightarrow$ **R** (vagy $\mathbf{R}^{k\times k}$) *veszteségfüggvény*, azaz W(T, t) a döntésünk nyomán támadt veszteség, ha t az igazi paraméterérték és g(t)-t T-vel becsüljük. Fel szokták tenni, hogy W nemnegatív (ill. pozitív szemidefinit), W(g(t), t) = 0, és W elsõ változójában konvex. Gyakran W(t) = h(T - g(t)) alakú. *Négyzetes veszteségfüggvényrõl* akkor beszélünk, ha vagy $h(x) = |x|^2$, vagy pedig $h(x) = xx^T$ (az elõbbi nem más, mint az utóbbi nyoma, azaz trace-e). Ez valóban konvex, mert tetszõleges 0 < a < 1 esetén

$$a\, h(x) + (1-a)\, h(y) - h(ax + (1-a)y) = a(1-a)\, h(x-y),$$

ami pozitív szemidefinit. Mi a továbbiakban mindig a négyzetes veszteségfüggvényt használjuk.
A W veszteségfüggvény mellett a T becslés $R_T(t)$ *rizikófüggvénye* a veszteség várható értéke (mint a t paraméter függvénye). Speciálisan a négyzetes rizikó: $R_T(t) = Var_t(T) + b_T(t)b_T(t)^T$. Torzítatlan becslések négyzetes rizikója tehát a kovarianciamátrixuk (szórásnégyzetük).
A paraméter g(t) függvényének két becslése, T_1 és T_2 közül T_1 (egyenletesen) *jobb* (azaz nem rosszabb), mint T_2, ha T_1 rizikófüggvénye sehol sem nagyobb, mint T_2-é (mátrixok esetén a megfelelõ különbségnek kell pozitív szemidefinitnek lennie).

4. tétel

- Egy ***D*** becslésosztályban T *megengedhetõ* (*admisszibilis*), ha nincs olyan becslés D-ben, amely határozottan jobb nála (jobb, és a két rizikófüggvény nem mindenütt egyezik meg). T *optimális,* ha minden más ***D***-beli becslésnél jobb.Ha ***D*** a torzítatlan becslések osztálya és négyzetes veszteségfüggvényt használunk, "jobb" helyett "hatásosabb", "optimális" helyett "hatásos (efficiens)" a szóhasználat.
- ***D***-ben T *minimax* becslés, ha minimalizálja a rizikófüggvény supremumát (ehhez persze a skalár értékû veszteségfüggvényt használjuk).

Állítás

Az optimális becslés (ha létezik), megengedhetõ és minimax. Ha ***D*** konvex, akkor az opimális becslés (ha létezik) egyértelmû (persze ***P***-mm.).

- Legyen a paraméter g(t) függvényének az n elemû mintából kiszámolt becslése T_n. A (T_n) sorozat *konzisztens,* ha minden t esetén $T_n \rightarrow g(t)$ sztochasztikusan, ill. *erõsen konzisztens,* ha minden t esetén $T_n \rightarrow g(t)$ P_t-mm. Világos, hogy ha a T_n-hez tartozó R_n rizikófüggvény-sorozat pontonként 0-hoz tart, akkor (T_n) konzisztens.

Tétel (Cramér-Rao egyenlõtlenség)

Tegyük fel, hogy a statisztikai mezõre teljesül (R). Legyen T lokálisan korlátos L_2-normájú torzítatlan becslése a paraméter g(t) függvényének (Tudjuk, hogy ekkor g folytonosan differenciálható; jelölje G(t) = dg(t)/dt.) Ekkor

$$\mathrm{Var}_t(T) \geq G(t)\, I(t)^{-1} G(t)^T .$$

Speciálisan, ha T skalár értékû és p = 1, akkor $D_t^2(T) \geq [g'(t)]^2 / I(t)$.

A jobb oldalon álló kifejezés neve ***információs határ***. Bár, mint láttuk, a Fisher-információ nem, de az információs határ már invariáns az átparaméterezésre.

Ha T_n a g(t)-nek az n elemû mintából kiszámolt torzítatlan becslése, akkor a rá vonatkozó információs határ

$$n^{-1}\, G(t)\, I_1(t)^{-1} G(t)^T,$$

vagyis reguláris esetben a torzítatlan becslés szórásnégyzete legfeljebb 1/n nagyságrendben tarthat 0-hoz.
Cramér-Rao egyenlõtlenség torzított esetben: Legyen T a differenciálható g(t) függvény tetszõleges becslése. (Ilyenkor a b(t) torzítás is differenciálható, jelölje G(t) = dg(t)/dt, B(t) = db(t)/dt.) Ekkor

$$R_T(t) \geq [G(t) + B(t)]\, I(t)^{-1} [G(t) + B(t)]^T + b(t)b(t)^T.$$

Ennek segítségével megmutattuk, hogy indikátormintából a relatív gyakoriság megengedhetõ becslése a valószínûségnek az összes becslés között.