Estimatzaileen propietateak

Josemari Sarasola

Estatistika enpresara aplikatua

Zein estimatzaile parametro batentzat?

Parametro batentzat hainbat estimatzaile asma daiteke. Adibidez, populazio bateko μ batezbestekoa estimatzeko lagin batezbestekoa, baina baita ere datu handiena gehi txikiena zati bi, erdiko datu zenbaiten batezbestekoa eta abar erabil daitezke. Kontua da parametroaren estimatzaile ona hautatu behar dugula. Baina zer da estimatzaile "ona"? Estimatzerakoan propietate interesgarriak dituena.

Estimatzaileen propietateak

Propietate ohikoenak

Estimatzaile bati buruz aztertzen diren propietate ohikoenak honako hauek dira:

- alboragabetasuna,
- zehaztasuna,
- efizientzia,
- doitasuna,
- tinkotasuna,
- askitasuna.

Notazioari

Gogoratu parametro bat orokorrean θ izendatzen dela, eta horren estimatzaile bat $\hat{\theta}$.

Estimatzaileen propietateak: alboragabetasuna

Alboragabetasunaren alegoria: fusilaria eta itua

Estimatzaile bat fusilari baten moduan irudikatu dezakegu, eta parametroa jo nahi duen itua (puntu gorria). Fusilaria alboragabea izango da bere tiroek ituan zentratuta daudenean, itua bete-betean jotzen ez badute ere. Fusilaria begi-okerra bada edo fusila okerturik badago, bere tiroek ez dira egongo ituan zentratuta, eta orduan fusilaria alboratua dagoela esango dugu.

fusil alboragabe edo zentratua

fusil alboratu edo okerra (ezkerrera)

Estimatzaileen propietateak: alboragabetasuna

Estimatzaile alboragabea: Definizioa

$\hat{\theta}$ estimatzaile bat alboragabea (ingelesez, *unbiased*) dela esango dugu, bere batezbestekoa estimatu nahi den parametroaren balioarekin bat datorrenean:

$$E[\hat{\theta}] = \theta$$

Hau da, beste era batera esanda, estimatzaile alboragabeak estimatzaile zentratuak dira: haiekin parametroaren balio zehatza ez dugu asmatuko orokorrean, baina egiten ditugun estimazioak zentraturik daude parametroaren balio ezezagunaren inguruan.

Estimatzaileen propietateak: alboragabetasuna

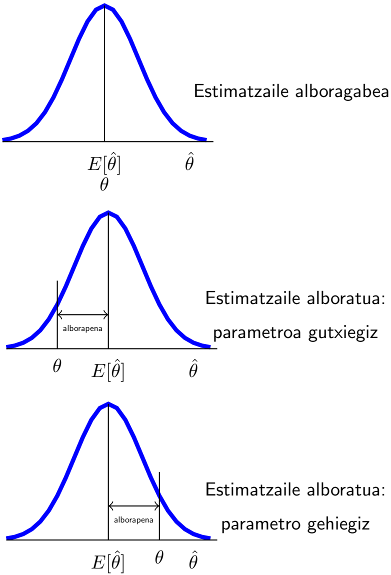
Definizioa: estimatzaile alboratua eta alborapena

$\hat{\theta}$ estimatzaile bat alboratua(ingelesez, *biased*) dela esango dugu, bere batezbestekoa estimatu nahi den parametroaren balioarekin bat ez datorrenean:

$$E[\hat{\theta}] \neq \theta$$

Hau da, estimatzailea alboratua da, okertuta dagoena batezbeste, estimatu nahi den parametroaren balioari buruz.
Alborapena (ing., *bias*; gazt. *sesgo*) $alb(\hat{\theta}) = E[\hat{\theta}] - \theta$ diferentzia da, eta estimazioak parametroaren balio ezezagunetik batezbestez zenbat desbideratuta (alboratuta) dauden eta norantz adierazten du.

Estimatzaileen propietateak: alboragabetasuna



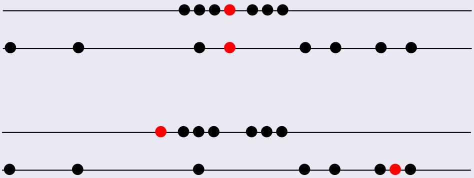
Estimatzaileen propietateak: alboragabetasuna

Estimatzaile asintokikoki alboragabeak

Batzuetan estimatzaileak alboragabeak ez badira ere, euren alborapena 0-rantz doa n lagin-tamaina handitu ahala. Kasu horietan, estimatzailea asintokikoki alboragabea dela esaten da. Nolabait esateko, estimatzaile bat alboragabea ez denean, asintotikoki alboragabea izatea eskatzen zaio.

Zehaztasunaren alegoria: fusilaria eta itua

Fusilaria zehatzagoa izango da bere tiroak, alboratuak ala ez, bilduagoak daudenean, sakabanatze txikiagoa dutenean alegia.



biak alboragabeak,
baina goiko fusila zehatzagoa da

biak alboratuak,
baina goiko fusila zehatzagoa da

Estimatzaileen propietateak: zehaztasuna

Nola neurtu zehaztasuna: bariantza

$\hat{\theta}$ estimatzaile baten zehaztasuna neurtzeko, horren bariantza kalkulatzen da:

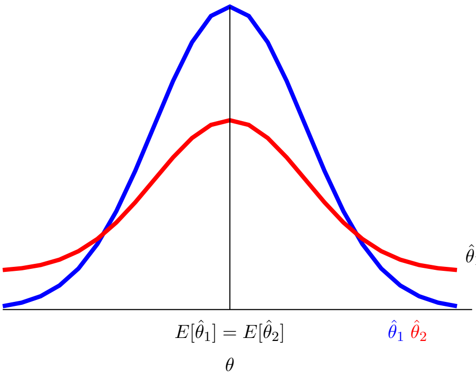
$$var(\hat{\theta})$$

Bariantza zenbat eta handiagoa izan, orduan eta zehaztasun txikiagoa dago.

Estimatzailearen errore estandarra

Estimatzailearen bariantzaren erroa, desbideratze estandarra alegia, estimatzailearen errore estandarra da, eta estimatzaile alboragabeen kasuan, estimazioak parametroaren baliotik batezbeste zenbat desbideratzen diren adierazten du.

Bi estimatzaile alboragabe, $\hat{\theta}_1$ eta $\hat{\theta}_2$. Zein da hobea?



$$var(\hat{\theta}_1) < var(\hat{\theta}_2)$$

$$\longrightarrow$$

$\hat{\theta}_1$ zehatzagoa da $\hat{\theta}_2$ baino (eta beraz, hobea)

Estimatzaileen propietateak: efizientzia

Efizientzia erlatiboa

$\hat{\theta}_1$ estimatzailearen efizientzia erlatiboa $\hat{\theta}_2$ estimatzaileari buruz honela definitzen da:

$$efe(\hat{\theta}_1, \hat{\theta}_2) = \frac{var[\hat{\theta}_2]}{var[\hat{\theta}_1]}$$

$\hat{\theta}_1$ efizienteagoa bada, $efe(\hat{\theta}_1, \hat{\theta}_2) > 1$ izango du.

Maiz, estimatzaileen bariantzak $\frac{k}{n}$ erakoak dira, k konstante izanik; horrelakoetan, $efe(\hat{\theta}_1, \hat{\theta}_2)$ efizientzia erlatiboak $\hat{\theta}_2$ estimatzailearen lagin-tamaina zenbat handitu behar den adierazten du, $\hat{\theta}_1$ estimatzailearen zehaztasun berdina izateko.

Cramér-Rao muga

Eredu parametriko batean (banaketa normalean, kasu), Cramér-Rao mugak parametro bati buruzko estimatzaile alboragabe ororen bariantzaren balio minimoa ezartzen du. Adibidez, banaketa normal batean μ parametroaren edozein estimatzaileren bariantzak gutxienez $\dfrac{\sigma^2}{n}$ balioa izango duela ezartzen Cramér-Rao mugak.

Estimatzaile efizienteak

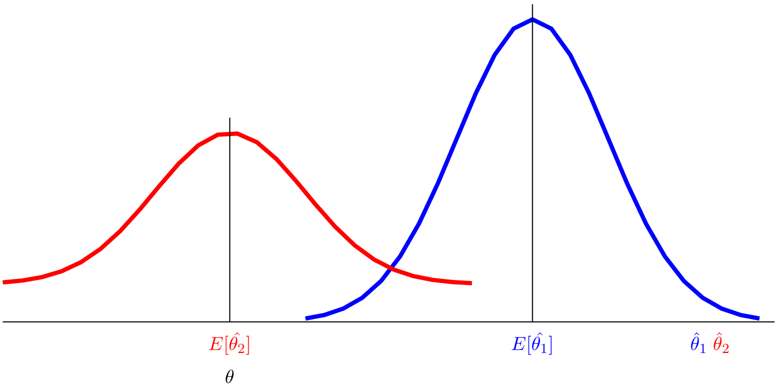
Estimatzaile alboragabe baten bariantzak parametro bati dagokion Cramér-Rao mugak ezartzen duen balioa hartzen badu, parametro horretarako lor daitekeen bariantza minimoa du, eta orduan estimatzaile *efizientea* dela esaten da. Adibidez, jakina da $var[\overline{x}] = \dfrac{\sigma^2}{n}$ dugula; beste alde batetik eredu normal bateko μ parametroaren Cramér-Rao muga $\dfrac{\sigma^2}{n}$ dela froga daiteke. Beraz, $\overline{x}$ estimatzailea estimatzaile efizientea da eredu normal bateko μ parametroari buruz.

Estimatzaileen propietateak: efizientzia

MVUE estimatzaileak

Cramér-Rao muga existitzen bada, ba al da beti estimatzaile efizienterik? Erantzuna ezezkoa da. Kasu horietarako, Cramér-Rao mugara heltzen ez garenean, MVUE (Minimum Variance Unbiased Estimator) delako estimatzaileak definitzen dira: bariantza txikieneko estimatzaile alboragabeak. MVUE estimatzailea beste estimatzaile alboragabe guztiak baino bariantza txikiagoa duena, hots efizienteagoa dena, da.
Aurreko definiziotik, ondorio hau erator daiteke: estimatzaile efizienteak MVUE estimatzaileak dira, baina badira MVUE estimatzaileak estimatzaile efiziente ez direnak (Cramér-Rao mugara heltzen ez direnean, hain zuzen).

Estimatzaileen propietateak: doitasuna



$\hat{\theta}_2$ alboragabea da, baina $\hat{\theta}_1$ zehatzagoa da. Zein zenbatesle hobesten da? Dilema ebazteko, doitasunaren irizpidea baliatu behar da.

Estimatzaileen propietateak: doitasuna

Doitasunak (ingelesez, *accuracy*) zenbateslearen balioak parametrora duen gertutasuna adierazten du, alborapena eta bariantza bateratuz. BEK edo batez besteko errore kuadratikoaren (ing., *MSE, Mean Square Error*) bitartez neurtzen da:

$$BEK[\hat{\theta}] = (E[\hat{\theta}] - \theta)^2 + var(\hat{\theta}) = alb(\hat{\theta})^2 + var(\hat{\theta})$$

Zenbatesle doiagoak batezbesteko errore kuadratiko txikiagoa dutenak dira.

Estimatzaileen propietateak: doitasuna

Zenbatesle doiaren bila

BEK estimatu nahi den θ parametroaren mendean izaten da, eta parametroak definizioz ezezagunak direnez gero, orokorrean ezin izaten da baieztatu estimatzaile baten BEK handiagoa den beste estimatzaile bat baino: θ parametroaren balio batzuetarako hala gertatuko da, eta beste batzuetarako ez. Nola aukeratu zenbatesle doiena kasu horietan?

Estimatzaileen propietateak: doitasuna

Estimatzaile menderatzaileak eta onartezinak

BEK irizpidea harturik, *parametroaren balio posible guztietarako* $\hat{\theta}_1$ estimatzaileak BEK txikiagoa badu $\hat{\theta}_2$ estimatzailea baino, $\hat{\theta}_1$ estimatzailea menderatzailea (ing., *dominating*) dela esaten da, eta $\hat{\theta}_2$ onartezina (ing., *inadmissible*).

Aztergai diren estimatzaileen artean, estimatzaile onartezinak baztertu behar dira.

Estimatzaileen propietateak: doitasuna

Estimatzaile onargarriak

BEK irizpidea harturik, *parametroaren balio posible batzuetarako* $\hat{\theta}_o$ estimatzaileak BEK txikiagoa (edo berdina) badu aztergai diren beste estimatzaile guztiek baino, parametroaren beste balio posible batzuetarako BEK handiagoa izan arren, $\hat{\theta}_o$ estimatzailea *onargarria* (ing., *admissible*) dela esaten da.

Estimatzaile onargarrien artean aukeratu behar da erabili beharrekoa.

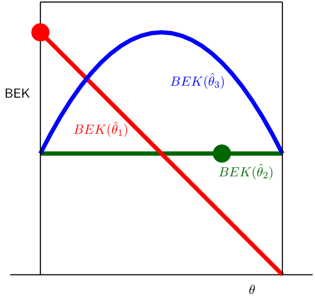
Doitasun irizpide bat: minimax

Minimax irizpidea

Estimatzaile onargarrien artean nola erabaki behar da erabili beharrekoa? Izan ere, parametroaren balio batzuetarako estimatzaile jakin bat izango da hobea, balio horietarako BEK txikiagoa duelako, baina beste batzuetarako beste bat.

Minimax irizpideak estimatzaile batek izan dezakeen BEK maximo edo handiena hartzen du erreferentzia gisa (badaezpada ere, parametroaren balioa ezagutzen ez dugunez) eta BEK maximo horietatik, txikiena duena (minimax: minimotu maximoak) aukeratu behar da.

Doitasun irizpide bat: minimax



Adibidez, irudian hiru estimatzaileren batez besteko errore kuadratikoak azaltzen dira, θ parametroak har ditzakeen balioen arabera. $\hat{\theta}_3$ estimatzaileak BEK balio handiagoak ditu $\hat{\theta}_2$ estimatzaileak baino, parametroaren balio guztietarako. Beraz, $\hat{\theta}_3$ estimatzaile onartezina da. Beste bi estimatzaileak, $\hat{\theta}_1$ eta $\hat{\theta}_2$, onargarriak dira, parametroaren balio batzuetarako BEK txikiena dutelako: minimax irizpidea erabiliz ordea, $\hat{\theta}_2$ doiena da, BEK maximo (puntu lodiz adierazita) txikiena duelako.

Estimatzaileen propietateak: konsistentzia

Estimatzaile konsistentea: definizioa

Estimatzaile bat konsistentea edo tinkoa dela esaten da, lagin tamaina handitu ahala, estimatu nahi den parametroaren baliora hurbiltzen denean. Zehatzago, estimatzailea konsistentea da parametroaren baliotik ϵ distantzia bat baino handiagoa desbideratzeko probabilitatea 0 denean, lagin tamaina aski handi baterako (labur esanda, estimatzailea parametroari buruzko probabilitatezko konbergentzia duenean. Praktikan, esan daiteke estimatzailea konsistentea izango dela alboragabea (edo asintokikoki alboragabea) denean eta bere bariantza 0-rantz doanean, n lagin-tamaina handitzean.

Estimatzaileen propietateak: askitasuna

Estimatzaile askia: definizioa

Estimatzaile bat askia dela esaten da, estimatzaileak laginak parametroari buruz jasotzen duen informazio guztia biltzen duenean, hots laginaren ezagutza zehatzak parametroari buruzko informaziorik ematen ez duenean. Definizio hau operazional bilakatzen da laginaren banaketa estimatzailearen balioaren baldintzapean parametroaren mendean ez dagoela frogatuz zein Fisher-Neyman faktorizazio-teoremaren bitartez. Ronald Fisher estatistikariak garatu zuen propietatea 1922 urtean.