Estimatzaileen lagin banaketak

Josemari Sarasola

Gizapedia

Lagin banaketak

Estimatzaileak

Parametroak (θ) kuantifikatzeko estimatzaileak ($\hat{\theta}$) erabiltzen dira. Estimatzaileak lagineko datuak erabiltzen dituzten formulak besterik ez dira. Adibidez, $\hat{\mu} = \overline{x}$ adierazpenak, μ populazioaren batezbestekoaren estimatzaile gisa ($\hat{\mu}$) lagin batezbestekoa erabiltzen dela adierazten du.

Lagin banaketak

Estimatzaileen lagin-banaketak

- Estimatzaileen balioak aldakorrak dira jasotzen diren datuen arabera, laginen arabera alegia.
- Laginak zorizkoak direnez (edo hala suposatzen dugu), estimatzaileek zorizko balioak hartzen dituzte, eta beraz probabilitate-banaketa izango dute.
- Zorizko balioak laginaren araberakoak direnez, estimatzailearen probabilitate-banaketari lagin-banaketa deitzen zaio.

Ikus ditzagun hurrengo gardenkietan zeintzu diren lagin-banaketa arruntenak.

Lagin batezbestekoa, eredu normala, σ ezaguna

$$\overline{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

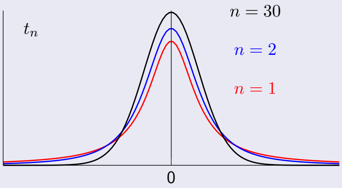
Lagin batezbestekoa, eredu normala, σ ezezaguna

$$t = \frac{\overline{x} - \mu}{\hat{s}/\sqrt{n}} \sim t_{n-1}$$

Gogoratu: $\hat{s} = \sqrt{\frac{\sum_i (x_i - \overline{x})^2}{n-1}}$.

Eranskina: Student-en t banaketa

- Honela idazten da labur: $t \sim t_n$.
- Askatasun-gradu kopurua deritzon n du parametro bakarra eta zenbaki naturala (1,2,...) izan behar da.
- n zenbat eta txikiagoa den, banaketa normal estandarrak baino mutur orduan eta astunagoak ditu. $n > 30$ denean, banaketa normal estandarraren ia berdina da. Hura bezalaxe, simetrikoa da $x = 0$ ardatzari buruz.



Eranskina: Student-en t banaketa

- Student-en t banaketaren balioak taularatuta daude $n \leq 30$ balioetarako.
- Taulak azpitik probabilitate zehatzak uzten dituzten balioak ematen ditu. 0.5eko beherako probabilitateetarako simetriaren propietatea erabiltzen da.
- Adibidez:
  - $t \sim t_4;\ P[t < t_0] = 0.99 \rightarrow t_0 = 3.75$
  - $t \sim t_7;\ P[t < t_0] = 0.1 \rightarrow t_0 = -1.42$
- Askatasun-graduak 30 baino gehiago direnean, Student t banakuntza N(0,1) banaketa normal estandar bilakatzen da.
- William Sealy Gosset kimikariak aurkitu zuen lagin txikien azterketan, garagardoen propietateen ikerketan. Ikerketa horiek *Student* ezizenarekin argitaratu zituen 1908 urtean eta hortik datorkio izena.

Estimatzaileen lagin banaketak

Lagin batezbestekoa, eredu ez normala, σ ezaguna

Lagin-tamaina handia $(n > 30)$ izan behar da:

$$\overline{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Estimatzaileen lagin banaketak

Lagin batezbestekoa, eredu ez normala, σ ezezaguna

Lagin-tamaina handiaren kasuan $(n > 30)$:

$$\overline{x} \sim N(\mu, \frac{\hat{s}}{\sqrt{n}})$$

Lagin batezbestekoa: laburpena

	σ ezaguna	σ ezezaguna
Pop. normala	$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$t = \dfrac{\overline{x} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$
Pop. ez-normala, $n > 30$	$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$\overline{x} \sim N\left(\mu, \frac{\hat{s}}{\sqrt{n}}\right)$

$\hat{p}$ lagin proportzioa

Adibidez, 20 pieza jasota lagin batean, 4 akastun badira, akastunen lagin proportzioa 4/20=0.2 da. p populazioko proportzioa zenbatesteko erabiltzen da. Horren lagin banaketa hau da, $n > 30$ lagin-tamainetarako:

$$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

s^2 lagin bariantza, eredu normala

$$\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

Gogoratu:

- $s = \sqrt{\frac{\sum_i (x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_i x_i^2}{n} - \overline{x}^2}$
- $\hat{s} = \sqrt{\frac{\sum_i (x_i - \overline{x})^2}{n-1}}$
- $\hat{s}^2 = \frac{n}{n-1} s^2$
- $s^2 = \frac{n-1}{n} \hat{s}^2$