Banaketa teorikoak: t, khi karratu eta F

Josemari Sarasola

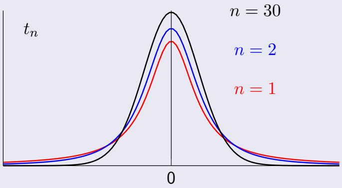
Estatistika enpresara aplikatua

Gizapedia

Lagin bateko datuekin batezbestekoak, bariantzak eta bestelako estimatzaileak kalkulatzen dira populazio bateko parametroak estimatzeko. Datuak zorizkoak direnez, estimatzaile horiek ere zorizkoak dira, eta banaketa teoriko jakin batzuei jarraitzen diete. Ondoren, banaketa teoriko horietatik arruntenak ikasi behar ditugu. Horien ezaugarriak eta haietarako taulak nola erabili ikasi behar ditugu.

Student-en t banaketa

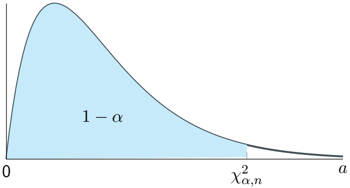
- Kanpai itxurako banaketa da, simetrikoa 0 balioaren inguruan, eta banaketa normalaren oso antzekoa, baina mutur astunxeagoekin.
- Askatasun-gradu kopurua deritzon n du parametro bakarra eta zenbaki naturala (1,2,...) izan behar da.
- Honela idazten da labur: $t \sim t_n$.
- $n > 30$ denean, banaketa normal estandarraren ia berdina da.



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:
  - $t \sim t_4;\ P[t < t_0] = 0.99 \rightarrow t_0 = 3.75 \rightarrow t_{4,0.01} = 3.75$
  - $t \sim t_7;\ P[t < t_0] = 0.1 \rightarrow t_0 = -1.42 \rightarrow t_{7,0.9} = -1.42$
- 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.

- χ^2_n (khi-karratu) banaketak, n parametro bakarra du, (*askatasun-gradu kopurua* izenekoa eta zenbaki naturala (1, 2, 3, ...) izan behar dena
- Balio positiboak soilik hartzen ditu eta eskubirantz alboratua da beti.



Bere balioak taularatuta daude, $1-\alpha$ azpiko probabilitate zehatzetarako. Adibidez,

- $\chi^2_{0.01,4} = 13.3$
- $\chi^2_{0.25,2} = 2.77$

$X \sim \chi^2_n$ banaketan n handia denean , $\frac{X-n}{\sqrt{2n}}$ aldagaiak banaketa normal estandarrerako konberjentzia erakusten du (nahiz eta konberjentzia Student-en t banaketan baino motelago gertatzen den). Adibidez, $\chi^2_{0.01,80}$ kalkulatu nahi bada,

$$z_{0.01} = 2.32 \rightarrow \chi^2_{0.01,80} = 2.32 \times \sqrt{80 \times 2} + 80 = 109.34$$

- X_1 eta X_2 aldagaiak khi-karratu banaketaren araberakoak badira, n_1 eta n_2 askatasun-gradu kopuruekin, orduan
- $F = \frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}}$ aldagaiak banaketa jakin bati jarraitzen dio, Snedecor edo Fisher-Snedecor $\mathscr{F}$ izenekoa:

$$F \sim \mathscr{F}_{n_1,n_2}$$

- n_1 eta n_2 parametroen arabera, pertzentil erabilienak taularatuta ditu.
- Alderanzketa-propietatea: $X \sim \mathscr{F}_{n_1,n_2} \rightarrow \frac{1}{X} \sim \mathscr{F}_{n_2,n_1}$
- Tauletan agertzen ez diren balioak emateko baliatzen da. Adibidez, $X \sim \mathscr{F}_{1,5} \rightarrow Y = \frac{1}{X} \sim \mathscr{F}_{5,1}$

$$P[X>10.01]=0.025 \rightarrow P[Y<\frac{1}{10.01}]=P[Y<0.099]=0.025 \rightarrow P[Y>0.099]=0.975$$