Banaketa hipergeometrikoa

Josemari Sarasola

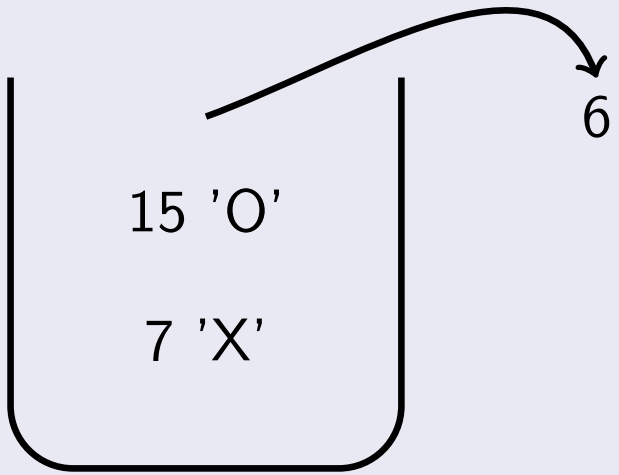
Estatistika enpresara aplikatua

Gizapedia

Ontzi problemak

Ontzi batean 15 pieza akasgabe eta 7 pieza akastun daude. 6 pieza ateratzen dira zoriz. Zenbat da haien artean 2 akastun izateko probabilitatea?

$$P[X=2] = \frac{\binom{7}{2}\binom{15}{4}}{\binom{22}{6}}$$



Banaketa hipergeometrikoa

Koefiziente binomialak

- $n = \dfrac{n!}{k!(n-k)!}$

Koefiziente jakingarri batzuk

- $\binom{n}{0} = 1$
- $\binom{n}{1} = n$
- $\binom{n}{n} = 1$

Koefiziente binomialetarako lasterbidea

$$\binom{8}{3} = \frac{8 \times 7 \times 6}{3!} = \frac{8 \times 7 \times 6}{6} = 8 \times 7 = 56$$

Hots, goiko zenbakia beheko zenbakia adina aldiz bidertzen da beheraka, eta ondoren beheko zenbakiaren faktorialarekin zatitu.

Ontzi problemara itzuliz

Gogoratu, [15 O, 7 X] ontzi batetik 6 aterata, 2 X izateko probabilitatea:

$$P[X=2] = \frac{\binom{7}{2}\binom{15}{4}}{\binom{22}{6}} = \frac{\frac{7!}{2!5!} \times \frac{15!}{4!11!}}{\frac{22!}{6!16!}}$$

$$= \frac{\frac{7 \times 6}{2!} \times \frac{15 \times 14 \times 13 \times 11}{4!}}{\frac{22 \times 21 \times 20 \times 19 \times 18 \times 17}{6!}}$$

Probabilitateak beste era batera kalkulatuz

Gogoratu, [15 O, 7 X] ontzi batetik 6 aterata, 2 X izateko probabilitatea:

$$P[X=2] = P[XXOOOOeo]$$
$$= \frac{7}{22} \times \frac{6}{21} \times \frac{15}{20} \times \frac{14}{19} \times \frac{13}{18} \times \frac{12}{17} \times \frac{6!}{2!4!}$$

eo horrekin 'edozein ordenatan' adierazi nahi da: 2 X eta 4 O ordenatzeko modu guztiak probabilitate berekoak direnez, aski da haietako bat hartu (XXOOOO) eta permutazio kopuruekin, haiek ordenatzeko era kopuruarekin alegia, bidertzea. Kasu honetan XXOOOO elementuen permutazioak $\frac{6!}{2!4!}$ dira.

Probabilitateen kalkulua R softwarean

```
>dhyper(2,7,15,6) #[15 0, 7 X] ontzi batetik 6
aterata, 2 X izateko probabilitatea
>phyper(2,7,15,6) #[15 0, 7 X] ontzi batetik 6
aterata, 2 X edo gutxiago izateko probabilitatea
>x=0:6
>dhyper(x,7,15,6) #[15 0, 7 X] ontzi batetik 6
aterata, x X izateko probabilitateak, x=0,1,2,3,4,5,6
balioetarako
```

Ontzi problemetatik itzulerarik gabeko laginketara

Ontzi problemak populazio finitu batetik lagin bat erauzteko problemara heda daitezke, zorizko laginketa itzulerarik gabe egiten denean. Adibidez, 400 emakume eta 600 gizon dituen 1000 tamainako populazio batetik, 100 tamainako lagin bat aukeratuta, 50 emakume eta 50 gizon suertatzeko probabilitatea hau da:

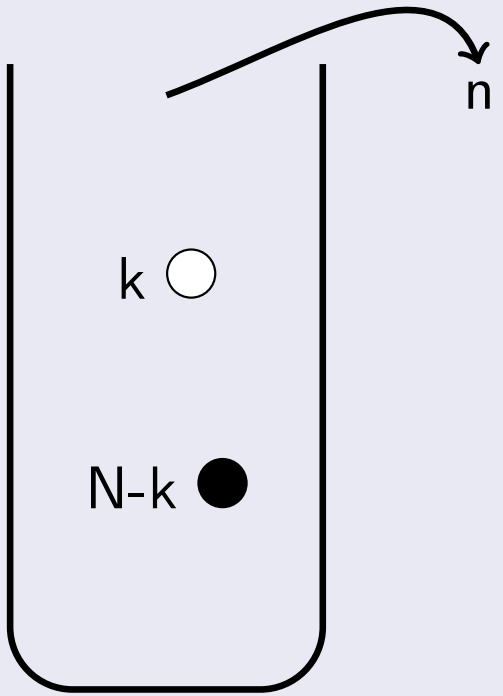
$$P[X=2] = \frac{\binom{400}{50}\binom{600}{50}}{\binom{1000}{100}}$$

Probabilitate funtzioa

Banaketa hipergeometrikoak N tamainako populazio dikotomiko batetik n tamainako lagin bat aukeratuta itzulerarik gabe (edota denak batera hartuta, finean prozesu berdina baita), haien artean arrakastatzat (○) harturiko k elementutik x elementu suertatzeko probabilitatea ematen du:

$$P[X = x; N, k, n] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$x = 0, 1, 2, \ldots, n$$



Simetriak

Ontzi batean N pilota izanda, k zuri eta $N-k$ beltz, n pilota aterata:

- $P[X=x;N,k,n]=P[X=n-x;N,N-k,n]$ (adibidez, 10 pilota aterata, 4 pilota beltz eta 6 pilota zuri ateratzeko probabilitateak berdinak dira).
- $P[X=x;N,k,n]=P[X=k-x;N,k,N-n]$ (adibidez, 30 pilotako ontzi batetik - 18 beltz, 12 zuri - 10 pilota aterata, 4 pilota beltz atera, eta ontzian geratu ziren 20 piloten artean 14 pilota beltz izateko probabilitateak berdinak dira).
- $P[X=x;N,k,n]=P[X=x;N,n,k]$ (30 pilotako ontzi batetik - 18 beltz, 12 zuri - 10 pilota aterata, 4 pilota beltz ateratzeko probabilitatea eta 18 beltzetan 4 pilota atera direnen artean izateko probabilitateak berdinak dira).
- $P[X=x;N,k,n]=P[X=N-n-k+x;N,N-k,N-n]$, lehenengo simetriari bigarrena aplikatuz (adibidez, 30 pilotako ontzi batetik - 18 beltz, 12 zuri - 10 pilota aterata, 7 pilota beltz atera, eta ontzian geratu ziren 20 piloten artean 9 pilota zuri izateko probabilitateak berdinak dira).

Erreferentzia: https://statistics.stanford.edu/research/tables-hypergeometric-probability-distribution

Balio posibleak

- Banaketa hipergeometrikoan $\bigcirc$ elementuen kopurua orokorrean 0-tik n-ra bitartekoa den arren, egoera berezietan balio posibleen tarte hau mugatuagoa da.
- Adibidez, 3 pilota zuri eta 4 pilota beltz dituen ontzi batetik 5 pilota ateratzen badira zoriz, pilota zurien kopurua 1 da gutxienez, eta 3 gehienez. Banaketa hipergeometrikoan balio posibleak, probabilitate positiboarekin, 1, 2 eta 3 izango dira (eta ez 0, 1, 2, 3, 4 eta 5).
- Zehatzago, beraz, banaketa hipergeometrikoan balio posibleak honako hauek dira:

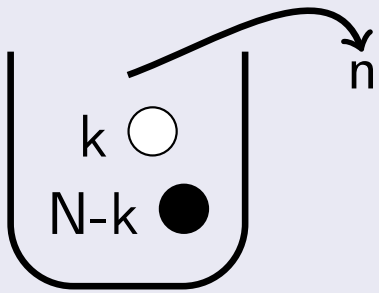
$$\max(0, n + k - N), \ldots, \min(k, n)$$

Adierazpena eta parametroak

$$X \sim H(k, N, n) \begin{cases} \mu = \dfrac{nk}{N} \\ \sigma^2 = \dfrac{nk}{N} \times \dfrac{N-k}{N} \times \dfrac{N-n}{N-1} \\ \text{Mo} = \left\lfloor \dfrac{(n+1)(K+1)}{N+2} \right\rfloor \end{cases}$$

Banaketa binomialarekin lotuz

- Banaketa hipergeometrikoa bezala, banaketa binomialak ere populazio dikotomiko batetik erauzitako lagin batean elementu kopuru jakin bat suertatzeko probabilitatea ematen du, baina hipergeometrikoan laginketa itzulerarik gabe egiten den bitartean, binomialean itzuleraz egiten da.



- Honela kalkulatzen da, itzuleraz, eta beraz binomialarekin, x arrakasta izateko probabilitatea:

$$P[X = x] = \left(\frac{k}{N}\right)^{x} \times \left(\frac{N-k}{N}\right)^{n-x} \times \frac{n!}{x!(n-x)!}$$

Banaketa binomialarekin lotuz

- Banaketa binomialean erauzketak independenteak diren bitartean, probabilitatea konstate mantentzen baitira aukeraketa guztietan, banaketa hipergeometrikoan dependentzia dago, aldi bakoitzean elementu bat gutxiago dagoelako populazioan, mota batekoa edo bestekoa, aurretik zenbat atera diren mota horietakoak:

$$P[X=x] = \overbrace{\frac{k}{N} \times \frac{k-1}{N-1} \times \cdots \times \frac{k-(x-1)}{N-(x-1)}}^{x\ arrakasta} \times \underbrace{\frac{N-k}{N-x} \times \cdots \times \frac{N-(n-x)-1}{N-(n-1)}}_{n-x\ porrot} \underbrace{\frac{n!}{x!(n-x)!}}_{edozein\ ordenatan}$$

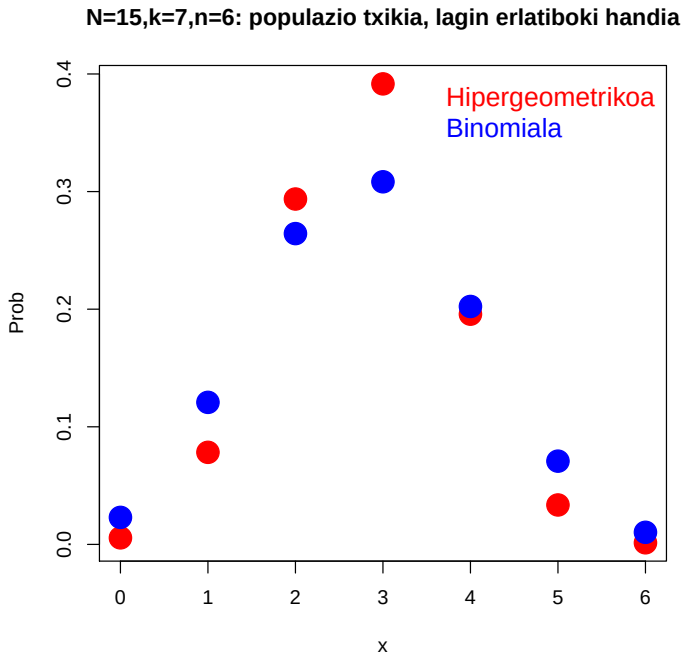
Banaketa binomialarekin lotuz

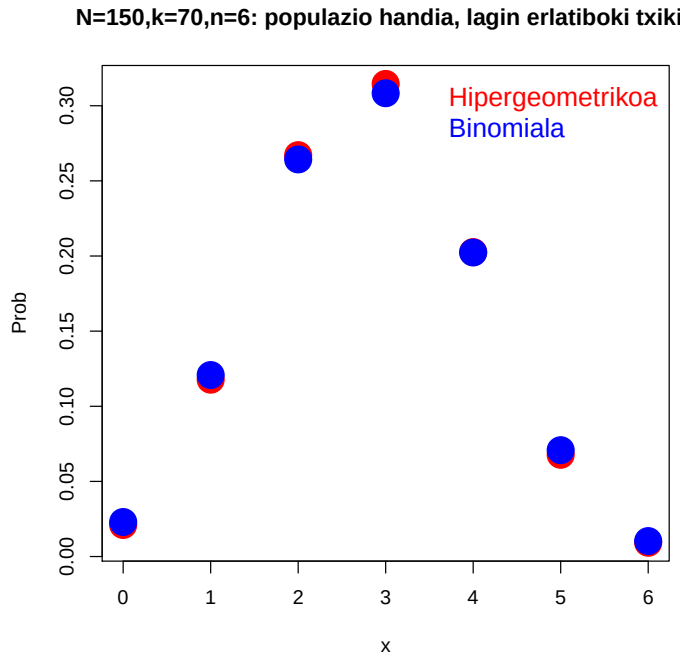
Banaketa	Itxaropena	Bariantza
$B(n, p = \frac{k}{N})$	$np = \frac{nk}{N}$	$npq = \frac{nk}{N} \times \frac{N-k}{N}$
$H(k, N, n)$	$\frac{nk}{N}$	$\frac{nk}{N} \times \frac{N-k}{N} \times \frac{N-n}{N-1}$

- Itxaropen berdina dituzte banaketa binomialak eta hipergeometrikoak: n pieza aukeratuta arrakasta kopuru berdina espero da bietan; baina hipergeometrikoan bariantza, eta beraz sakabanatzea, txikiagoa da ($(N-n)/(N-1) < 1$ delako).
- Bariantza txikiago horren azalpena intuitiboa ere bada: banaketa hipergeometrikoan, elementuak aukeratu ahala, populazio edo ontzia gero eta txikiago egiten da, eta ziurtasun handiagoa izango da aukeratu den elementu motari buruz.

Banaketa binomialarekin lotuz

- Finean, banaketa hipergeometrikoaren bariantzaren $(N-n)/(N-1)$ faktorearen arabera dira bi banaketak ezberdinak:
  - N populazioaren tamaina oso handia denean, faktorea 1-etik gertu izango da, eta bi banaketak antzekoak dira;
  - izan ere, binomialean banakako $\frac{k}{N}$ probabilitateen zatitzaileak beti N diren bitartean, hipergeometrikoan elementuak atera ahala N, $N-1$, ... zatitzaileen segidak pixkanaka-pixkanaka egiten du beheraka populazio handi batean.
  - n lagin-tamaina txikia denean ere, N populazioaren tamainarekiko, bi banaketak antzekoak izango dira.
  - laburrean, bi banaketak oso antzekoak izango dira populazio handietan lagin txikiak aukeratzen direnean





Irudian ikus daitekeenez, populazioa handia eta lagina erlatiboki txikia direnean banaketa hipergeometrikoa eta binomiala antzekoak dira.

Banaketa binomialarekin lotuz

- Beraz, populazioa handia eta lagin tamaina txikia direnean, itzulerarik gabeko zorizko laginketa egin arren, ohikoa da banaketa binomiala erabiltzea oinarri eta eredu moduan, aukeratutako elementuen arteko independentziak asko errazten baitu laginketa problemen ebazpena.
- Populazio finitu txikien eta lagin tamaina handien kasuan ere (laginek populazioaren %5-10-etik gora osatzen dutenean, zehatzago), independentziatik eratorritako formulak aplikatzen dira, eta haiei banaketa hipergeometrikoaren bariantzaren $(N-n)/(N-1) < 1$ faktore zuzentzailea aplikatu ohi zaie.