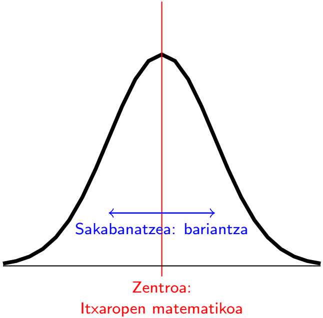
Probabilitate banaketen ezaugarriak

Josemari Sarasola

Gizapedia

Estatistika enpresara aplikatua

Datu-multzoetan bezalaxe, probabilitate banakuntzetan ere zentroa, sakabanatzea eta beste ezaugarri estatistiko batzuk azter daitezke. Ikasgai honetan, zentroa ikasiko dugu, itxaropen edo esperantza matematikoaren bitartez, eta sakabanatzea, bariantzaren bitartez.



Itxaropen matematikoa

Itxaropen matematikoa, ingelesez *expected value*, probabilitate banakuntzetako zentroa adierazten du. Batzuetan, *batezbesteko* ere deitzen zaio, nahiz eta datuekin kalkulatutako batezbestekotik bereizi behar den. Honela izendatu eta kalkulatzen da (μ: mu):

- banakuntza diskretuetarako,

$$\mu = E[X] = \sum_{x \in \Omega} x \cdot p(x)$$

- banakuntza jarraituetarako,

$$\mu = E[X] = \int_{\Omega} x f(x) dx$$

Dado bat botata, zenbat da puntu kopuruaren itxaropena?

x	1	2	3	4	5	6	
p(x)	1/6	1/6	1/6	1/6	1/6	1/6	1
xp(x)							$\mu = \frac{21}{6} = 3.5$

Beraz, dadoa botata, batezbeste 3.5eko puntuazioa espero behar dugu.

Orain, berriz, dadoa aldi batzuetan bota eta puntu hauek izan dira: 4, 5 eta 6. Puntuen batezbestekoa $\overline{x}$ =(4+5+6)/3=5 da.

Itxaropen matematikoa eta batezbestekoa

Zergatik dira ezberdinak μ eta $\overline{x}$? Kontzeptu bera jaso arren (zentroa), ikuspuntu eta egoera ezberdinetatik kalkulatzen direlako.

μ, itxaropena	$\overline{x}$, batezbesteko aritmetikoa
• finkoa da	• aldakorra da, datuak zein diren
• probabilitate banaketatik	• datuetatik
• ideala	• konkretua
• a priori (esperantza)	• a posteriori, datuak jasota
• epe luzera hurbiltzen zaio $\overline{x}$	• datu gutxirekin (epe laburrera)
	μ-ren oso ezberdina izaten da
• parametroa izaten da	• parametroaren estimatzailetzat hartzen da
• ezezaguna da askotan	• beti kalkula daiteke, datuak harturik

Itxaropenaren propietateak

(1) *zorizko aldagaien baturari buruz,*

$$X_1, X_2, \ldots, X_n \ z.a.$$

$$\mathbf{X} = X_1 + X_2 + \ldots + X_n \rightarrow E[\mathbf{X}] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

Adibidez, 10 dadoen puntuazio totalaren itxaropena 10 dadoen itxaropenen batura da: $3.5 + \cdots + 3.5 = 35$

(2) *aldagai aldaketa linealari buruz,*

$$X \ z.a.$$

$$Y = a + bX \rightarrow E[Y] = a + bE[X]$$

Adibidez, dadoan jolasteagatik puntu bakoitzeko 10€ eta 20€ finko ematen badidate:

$dirua = 20 + 10X \rightarrow E[dirua] = 20 + 10E[X] = 20 + 10 \times 3.5 = 55$

Bariantza gehien erabiltzen den sakabanatze-neurria da. Honela izendatzen da (σ: sigma minuskula):

$$\sigma_X^2 = var[X]$$

Honela kalkula daiteke modu erosoan:

$$\sigma_X^2 = \alpha_2 - \alpha_1^2$$

non α_2 honela kalkulatzen den:

- diskretuan: $\alpha_2 = \sum_\Omega x^2 p(x)$
- jarraituan: $\alpha_2 = \int_\Omega x^2 f(x) dx$

Sakabanatzea: bariantza

- Bereizi behar dira σ_X^2 (probabilitate-banakuntzen bariantza) eta s_X^2 (datuen bariantza).
- Oro har, probabilitate-banakuntzetako neurriak letra grekoz, eta datuenak letra latinez.
- Azkenik, *desbideratze estandarra* bariantzaren erroa da:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Bariantzaren propietateak

(1) *zorizko aldagaien baturari buruz,*

$$X_1, X_2, \ldots, X_n \ z.a.$$

beraien artean elkarrekiko independenteak,

$$\mathbf{X} = X_1 + X_2 + \ldots + X_n \rightarrow \sigma^2_{\mathbf{X}} = \sigma^2_{X_1} + \sigma^2_{X_2} + \ldots + \sigma^2_{X_n}$$

(2) *aldagai aldaketa linealari buruz,*

$$X \ z.a.$$

$$Y = a + bX \rightarrow \sigma^2_Y = b^2 \sigma^2_X$$

Itxaropena, erabaki-irizpide gisa

Probabilitatea ala itxaropena?

A ekoizpen prozesu batean, akastuna izateko probabilitatea 0.1 da; B prozesu batean, berriz, 0.05. Zein da prozesu optimoa? Akastuna izateko probabilitate txikiena duena hartuko genuke informazio soil horrekin: B prozesua, alegia.

Itxaropena, erabaki-irizpide gisa

Probabilitatea ala itxaropena?

Informazio gehigarria eskuratzen dugu: 4 eurotan saltzen dira piezak. A prozesuko unitateko kostua 2 da; B prozesukoa, berriz, 3. Orain, prozesu bakoitzean espero den mozkina kalkula daiteke:

x_A	$p(X_A)$	$x_A p(X_A)$	x_B	$p(X_B)$	$x_B p(X_B)$
-2	0.1	-0.2	-3	0.05	-0.15
2	0.9	1.8	1	0.95	0.9
		1.6			0.8

Eta horrela, A prozesua litzateke hoberena ikuspuntu horretatik.

Itxaropena, erabaki-irizpide gisa

Probabilitatea ala itxaropena?

Zein da, beraz, irizpideak irizpide, aukera onena? Irudi luke itxaropenaren irizpidea aberatsagoa dela, informazio handiagoa barneratzen duelako. Baina ez da horrela; besterik gabe, irizpide ezberdinak dira, esanahi ezberdinekoak, eta guk hautatu beharko dugu zein nahi dugun garatu, gure helburuen arabera.

Itxaropena, erabaki-irizpide gisa

Epe laburrera eta epe luzera

Txanpona botatzeko bi joko aukeran, A eta B. Zein nahiago duzu epe luzera? Eta epe laburrera, behin bakarrik jokatzeko zehazkiago?

x_A	$p(X_A)$	$x_A p(X_A)$	x_B	$p(X_B)$	$x_B p(X_B)$
-10	0.5	-5	-1000	0.5	-500
20	0.5	10	3000	0.5	1500
		5			1000

Epe laburrera eta epe luzera

Epe luzera itxaropenari bakarrik begiratu behar diogu; hartara, argi dago B jokoa hobetsi beharko genukeela, batezbestez epe luzera irabazi handiagoak ematen dituelako. Baina nor ausartzen da joko hori behin bakarrik jokatzera? Epe laburrera gauzak ez daude hain garbi. Ikus dezagun:

Epe laburrera eta epe luzera

- Epe laburrera itxaropenaz gainera (zenbat eta handiagoa, hainbat hobe), arriskua ere (zenbat eta handiagoa, hainbat eta okerrago) ebaluatu egin behar da.
- Arriskua honela definitu daiteke: emaitzak zenbateraino diren gorabeheratsuak.
- B jokoa arriskutsuagoa da, beraz.
- Arriskua saihestuz, arriskuaren aurka alegia, (gazteleraz, con aversión al riesgo) jokatu edo erabaki ohi da.
- Finantzetan, bereziki burtsan, arriskuari *hegazkortasuna* deritzo (ing., volatility).

Itxaropena, erabaki-irizpide gisa

Epe laburrera eta epe luzera

Arriskua bariantzaren bitartez neurtzen da sarri. Zenbat eta bariantza handiagoa, irabaziak gero eta aldakorragoak, eta beraz arriskua (irabazi oso handiak eta oso txikiak edo galerak izateko aukera) orduan eta handiagoa.
Beraz, **epe laburrera** bariantza txikiagoko aukerak hobesten dira epe laburrera, eta itxaropen handiagoko aukerak.

Utilitate-funtzioak

Epe laburrerra joko- edo inbertsio-aukera batek itxaropen handiagoa eta bariantza (eta beraz, arrisku) txikiagoa baditu beste batek baino, ez dago arazorik; horixe hobesten dugu. Zer gertatzen da itxaropen handiagoa (ona) eta bariantza handiagoa (txarra) edota bariantza txikiagoa (ona) eta itxaropen txikiagoa (txarra) baditu? Dilema sortzen zaigu horrelakoetan, eta erantzuna ez da garbia. Dilema ebatzi eta erantzun garbia emateko, *utilitate funtzioak* erabiltzen dira, μ eta σ harturik inbertsioaren *utilitate* edo onuragarritasunaren neurri bat ematen dutenak.

Itxaropena, erabaki-irizpide gisa

Utilitate-funtzioak

Ohiko utilitate-funtzioa hau da:

$$U = \frac{\mu}{\sigma}$$

Baina bestelakoa ere izan daiteke, baldintza edo axioma batzuk, *axiomatika* bat alegia, betetzen badu. Utilitate-funtzioen axiomatikaren baldintza hauek ikasiko ditugu:

- zenbat eta μ handiagoa, orduan eta utilitate handiagoa izan behar da;
- zenbat eta σ (edo σ^2, arriskua alegia) handiagoa, orduan eta utilitate txikiagoa izan behar da.

Itxaropena, erabaki-irizpide gisa

Utilitate-funtzioak

Dilemak ebazteko utilitate-funtzio egokia aukeratuta, utilitate handiena ematen duen aukera hobetsiko da.

Egunkari saltzailearen problema

Itxaropenaren bitartez ebazten den enpresa-problema ezaguna da, *revenue management* izeneko arloan. Egunkari-saltzaileak zenbat egunkari erosi behar dituen erabaki behar du, gero berriz ere saltzeko. 1 eurotan erosi, eta 4 eurotan saltzen ditu. Saldu gabeak bota egin behar ditu. Eskaria honela banatzen da:

x	1	2	3	4
$p(x)$	0.1	0.3	0.4	0.2

Zenbat egunkari erosi behar ditu aurrez?

Itxaropena, erabaki-irizpide gisa

Egunkari saltzailearen problema

$$s : salmentak; \;\; m : mozkinak$$

1 erosita

s	m	$p(m)$	$mp(m)$
1	4-1=3	1	3
			3

2 erosita

s	m	$p(m)$	$mp(m)$
1	$1 \times 4 - 2 \times 1 = 2$	0.1	0.2
2	$2 \times 4 - 2 \times 1 = 6$	0.9	5.4
			5.6

Egunkari saltzailearen problema

3 erosita

s	m	$p(m)$	$mp(m)$
1	$1 \times 4 - 3 \times 1 = 1$	0.1	0.1
2	$2 \times 4 - 3 \times 1 = 5$	0.3	1.5
3	$3 \times 4 - 3 \times 1 = 9$	0.6	5.4
			7

4 erosita

s	m	$p(m)$	$mp(m)$
1	$1 \times 4 - 4 \times 1 = 0$	0.1	0
2	$2 \times 4 - 4 \times 1 = 4$	0.3	1.2
3	$3 \times 4 - 4 \times 1 = 8$	0.4	3.2
4	$4 \times 4 - 4 \times 1 = 12$	0.2	2.4
			6.8

Egunkari saltzailearen problema

Beraz, erabaki optimoa 3 egunkari erostea da, horrela itxarondako mozkina maximotu egiten delako.

Itxaropena, erabaki-irizpide gisa

Egunkari saltzailearen problema

Egunkari saltzailearen probleman beste hainbat parametro azter daitezke:

- Galduko diren salmenten itxaropena
- Itxarondako salmentak
- Biltegian saldu gabe eta likidatu aurretik geratuko diren unitateen itxaropena
- Eskariaren asebetetze-ratioa (fill rate): itxarondako salmentak itxarondako eskariarekin zatituz
- Instock probabilitatea: eskaria betetzeko probabilitatea
- Stockout probabilitatea: eskaria galtzeko probabilitatea

Egunkari saltzailearen problema: eremu jarraitua

Sagar zukua saltzen duen enpresari batek zenbat zuku erosi behar duen erabaki behar du, gero berriz ere saltzeko. Litroa 1 eurotan erosi, eta 4 eurotan saltzen du. Saldu gabea bota egin behar du. Eskaria honela banatzen da:

$$f(x) = \frac{1}{4};\ 1 < x < 5$$

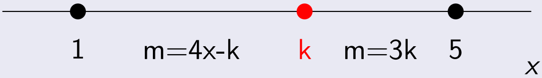
Zenbat zuku erosi behar du aurrez?

Itxaropena, erabaki-irizpide gisa

Egunkari saltzailearen problema: eremu jarraitua

m mozkina honela kalkulatzen da, k erosi beharreko kopurua izanik eta x eskaria izanik:

- $x < k \rightarrow m = 4x - 1 \times k = 4x - k$
- $x > k \rightarrow m = (4 - 1)k = 3k$



Itxaropena, erabaki-irizpide gisa

Egunkari saltzailearen problema: eremu jarraitua

Mozkinaren itxaropena honela kalkulatzen da:

$$
\begin{aligned}
E[m] &= \int_1^k \underbrace{(4x-k)}_{m} \overbrace{\frac{1}{4}}^{f(x)} \, dx + \int_k^5 \underbrace{3k}_{m} \overbrace{\frac{1}{4}}^{f(x)} \, dx \\
&= \int_1^k x - \frac{k}{4} dx + \int_k^5 \frac{3k}{4} dx \\
&= \left[\frac{x^2}{2} - \frac{kx}{4} \right]_1^k + \left[\frac{3kx}{4} \right]_k^5 = 4k - \frac{k^2}{2} - \frac{1}{2}
\end{aligned}
$$

Itxaropena, erabaki-irizpide gisa

Egunkari saltzailearen problema: eremu jarraitua

Mozkinaren itxaropena maximotzen duen k nahi dugunez, horri buruz deribatu, 0ra berdindu eta k bakantzen dugu:

$$\frac{dE[m]}{dk} = 0 \rightarrow 4 - \frac{2k}{2} = 0 \rightarrow k^* = 4\ l$$