Zorizko aldagaiak

Josemari Sarasola

Estatistika enpresara aplikatua

Gizapedia

Zorizko aldagaia eta probabilitate banaketa

Zorizko aldagaia zoriz zenbakizko balioak hartzen dituen aldagai bat da (ing. *random variable*, gazt., *variable aleatoria*). Adibidez, dado batean izandako puntuak, bikote batek izango duen haur kopurua eta ikasle batek azterketa batean lortuko duen nota.

Zorizko aldagai batek har ditzakeen balio guztiak horien probabilitateekin batera (egin klik) ematen direnean, zorizko aldagaiaren *probabilitate-banaketa* definitu dela esaten da (ing., *probability distribution*, gazt., *distribucion de probabilidad*).

Zorizko aldagaiak, eta beraz horien probabilitate banaketak, *diskretuak* nahiz *jarraituak* izan daitezke.

Zorizko aldagaia eta probabilitate banaketa

Adibidea: dadoa jaurtitzen denean, lortutako puntu kopuruari buruz,

zorizko aldagaia

x	$p(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$
	1

probabilitate banaketa

Zorizko aldagai diskretuak

Zorizko aldagai diskretuak balio bakanak edo isolatuak hartzen dituztenak dira (dado bateko puntu kopurua, bikote baten haur kopurua, adibidez). Bi eratara zehaz daiteke horren banaketa: *probabilitate-funtzioaren bitartez* nahiz *banaketa-funtzioaren bitartez*.

Probabilitate-funtzioak balio baten probabilitatea ematen du *zuzenean*. Taula batez zein funtzio matematiko batez definitua izan daiteke.

Adibidez, $P[X = x] = \dfrac{x+1}{10}; x = 0, 1, 2, 3$ probabilitate funtzioa da (X: aldagaia, x: balio konkretua). Taula moduan:

x	0	1	2	3
p(x)	$\dfrac{0+1}{10} = 0.1$	$\dfrac{1+1}{10} = 0.2$	$\dfrac{2+1}{10} = 0.3$	$\dfrac{3+1}{10} = 0.4$

Probabilitate funtzio batean, probabilitate guztien batura 1 izan behar da, noski.

Zorizko aldagai diskretuak

Banaketa-funtzioak probabilitate metatuak ematen ditu: x balio batetik beherako probabilitatea, alegia:

$$F(x) = P[X \leq x]$$

Taulaz zein funtzionalki adierazia izan daiteke. Adibidez:

$$F(x) = 1 - (1/2)^x; x = 1, 2, 3, \ldots$$

x	1	2	3	...
$F(x) = P[X \leq x]$	$1 - \left(\frac{1}{2}\right)^1 = 0.5$	$1 - \left(\frac{1}{2}\right)^2 = 0.75$	$1 - \left(\frac{1}{2}\right)^3 = 0.875$	...

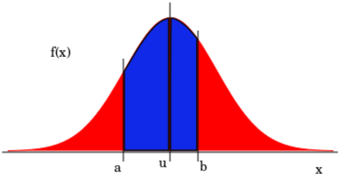
Propietate hauek ditu:

- probabilitateak goraka doaz, x-rekin batera, probabilitateak metatzen direnez;
- 0tik abiatu, eta 1 balioan bukatzen da.

Zorizko aldagai diskretuak: laburpena

- probabilitate-funtzioa $\rightarrow P[X = x] \rightarrow$ prob. sinpleak
- banaketa-funtzioa $\rightarrow F(x) = P[X \leq x] \rightarrow$ prob. metatuak

Zorizko aldagai jarraituak tarte batean edozein balio har dezaketenak dira (adib., pertsona baten altuera). *Dentsitate-funtzioaren bitartez* zein *banaketa-funtzioaren bitartez* zehaztu daiteke dagokion banaketa:



Goiko irudia *dentsitate-funtzioa* da, eta $f(x)$ adierazten da. Intuitiboki oso esanahi garbia du: ordenatua (funtzioaren altuera, alegia) altua den zatian, u inguruan alegia, probabilitatea handia da. Baino zehazkiago, tarte bateko probabilitatea tarte horretan funtzioak duen x ardatz bitarteko azalera dela esan behar da:

$$P[a < X < b] = \int_a^b f(x)dx$$

Zorizko aldagai jarraituak

Puntu bateko probabilitatea beti da 0, tarte bateko infinitu punturen artean puntu bat ez baita ezer ere, eta haren azpian ez dago azalerarik:

$$P[X = x] = 0$$

Beste alde batetik, dentsitate-funtzioek baldintza hauek bete behar dituzte:

1. Azpiko azalera osoa, zorizko aldagaia definitua dagoen balioetarako, balio posibleetarako alegia, 1 izan behar da:

$$\int_{\Omega} f(x)dx = 1$$

2. Balio posibleen tarte osoan zehar x ardatzaren gainetik dago:

$$f(x) \geq 0, \forall x \in \Omega$$

Banaketa-funtzioak balio batetik beherako probabilitatea ematen du (puntu hori bera sartzen den ala ez hutsala da, puntu bateko probabilitatea 0 denez):

$$F(x) = P[X < x] = \int_{inf}^{x} f(x)dx$$

Hala, x baliorainoko probabilitate metatua ematen duen funtzioa dela esan daiteke: x gehituz, dentsitate-funtzioko azpiko azalerak metatzen doa.

*inf: gerta daitekeen balio txikiena, sup: gerta daitekeen balio handiena

Beraz, banaketa-funtzioak *zuzenean* ematen du x azpiko probabilitatea, bertan dentsitate-funtzioan egin behar dugun integrala ja kalkulatuta dagoelako.

Baina banaketa funtzioarekin beste edozein tarteko probabilitatea ere kalkula daiteke, balioak banaketa-funtzioan behar bezala ordeztuz:

- $P[X{<}a]{=}F(x{=}a)$
- $P[X{>}b]{=}1{-}P[X{<}b]{=}1{-}F(x{=}b)$
- $P[a{<}X{<}b]{=}P[X{<}b]{-}P[X{<}a]{=}F(x{=}b){-}F(x{=}a)$

Zorizko aldagai jarraituak: banaketa-funtzioa

Banaketa-funtzioak propietate hauek betetzen ditu, beraz:

1. $F(x = inf) = 0$, balio posible txikienaren azpitik probabilitatea 0 delako.
2. $F(x = sup) = 1$, balio posible txikienaren gainetik probabilitatea 0 delako.
3. Gorakorra da, probabilitatea metatzen doalako, x-k aurrera egin ahala.
4. Eta beraz, aurrekoetatik, $0 \leq F(x) \leq 1$ betetzen duela esan daiteke, hau da, 0 eta 1 bitarteko balioak hartzen ditu beti.

Zorizko aldagai jarraituak: banaketa-funtzioa

Banaketa-funtzioak eta dentsitate-funtzioak elkarrekin duten erlazioa honelakoa da:

- dentsitate-funtziotik banaketa-funtzioa eskuratzeko, integratu behar da, dakigunez:

$$F(x) = \int_{inf}^{x} f(x)dx$$

- eta beraz, banaketa-funtziotik dentsitate-funtzioa eskuratzeko, deribatu behar da:

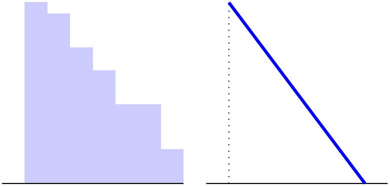
$$f(x) = \frac{dF(x)}{dx}$$

Parametro kontzeptua

- Probabilitate-banaketa batean hainbat balio ezberdin har ditzakeen konstante bat da, konstante zehazgabea alegia, letra batez adierazi ohi dena (ez x, hori zorizko aldagaia baita) eta balio horietako bakoitzeko banakuntza guztiz zehazten duena.
- Parametroa zehazten ez bada, probabilitate-banakuntza era generikoan definiturik geratzen da, zehaztu gabe, eta beraz probabilitateak ezin dira kalkulatu.
- Printzipioz, parametroak ezezagunak izaten dira probabilitate-banaketa batean.
- Garrantzitsuak dira estatistikan, helburua parametro ezezagun hori zenbatetsi edo estimatzea baita, lagin bateko datuetan oinarrituz, banakuntza guztiz zehazte aldera.

Banaketak eredu moduan

Har ditzagun zorizko aldagai bati esleitu diogun probabilitate-banaketa eta zorizko aldagai horri buruzko datu errealak:



$datu\ errealak$ $f(x)$

Probabilitate-banaketak horrelakoak izaten dira: matematikoak, eta beraz, oso perfektoak edo.
Errealitatea gorabeheratsuagoa da, datuetan ikusten dugunez.
Errealak al dira, beraz, probabilitate-banaketak?

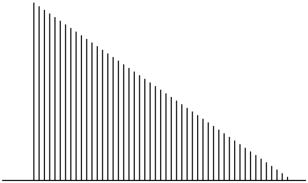
Banaketak eredu moduan

Erantzuna:

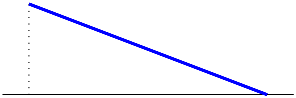
Ez dira guztiz errealak, baina zientzian ez dira bilatzen erabateko zehaztasuna eta xehetasun guztietara heltzea, errealitatearen hurbilketak edo sinplifikazioak baizik, *eredu edo modeloak* alegia, problemak era sinplean ebazteko planteatzen direnak. Izan ere, problemak errealitateko datu gordinekin lan egitea oso konplexua litzateke.

Horrexegatik idazten dugu problemetan *[...] banatzen dela uste da*, probabilitate-banaketak ereduak baitira, *usteak edo hipotesiak* finean, gero errealitateko datuekin kontrastatu behar direnak.

Zorizko aldagai diskretu batek balio asko hartzen dituenean, probabilitate-banaketa jarraitu baten bitartez hurbildu ohi da, laburtzeagatik (erosoagoa da kalkuluak egiteko):



$nota\ (diskretuak: 0, 0.1, 0.2, \ldots, 10)$



$nota\ (jarraitua: f(x); [0, 10])$

Nola kalkulatu $P[nota = 5]$?

- Diskretuan, besterik gabe taulako probabilitateari (zutabearen altuerari) erreparatu. Adibidez: $P[nota = 5] = 0.04$
- Jarraituan, 5 balioa puntua denez, $P[nota = 5] = 0$
- DESFASEA SORTZEN DA!
- Nola ebatzi desfase hori?

Probabilitate zuzena eremu diskretukoa da, noski. Izan ere, egon badaude ikasleak 5eko notarekin.

Nola lortu eremu jarraituan eremu diskretuko emaitza (gutxi gorabehera) eskuratzea?

$$P[X = 5] = P[4.95 < X < 5.05] = \int_{4.95}^{5.05} f(x)dx$$

Izan ere, eremu jarraituko notak nota diskretu gertuenera borobiltzen dira:

- $4.9765 \rightarrow 5$
- $5.0398 \rightarrow 5$
- baina, $4.9327 \rightarrow 4.9$
- baina, $5.0649 \rightarrow 5.1$

Jarraituzko hurbilketak

Beste adibide batzuk

- unibertsitate batean matrikulatutako ikasle kopuruari buruz:

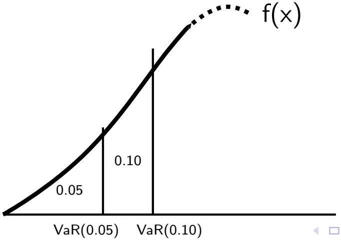
$$P[X = 1200] = P[1199.5 < X < 1200.5]$$

- 10 gramuko zehaztasuna duen balantza batean pisatzen den fruitu bati buruz:

$$P[X = 100] = P[95 < X < 105]$$

Arrisku-balioa (Value at Risk, VaR)

- Value at Risk (VaR) inbertitzaile zuhurrentzako adierazlea da, gerta daitekeen etekin txikia edo eskasa adierazten duena, okerrenera gerta daitekeen balioa (edo gutxiago) alegia, betiere aurrez zehaztutako probabilitate jakin baterako.
- Adibidez, aholkulariak esaten digu erosteko asmoa dugun akzioarekin 0.1eko probabilitateaz (hori berak finkatu du) irabaz dezakegula 2 baino gutxiago. Beraz, VaR(0.1)=2 dela adierazten digu finean.
- Baliatzen diren VaR ohikoenak 0.01 eta 0.05 probabilitateei dagozkienak dira.



Arrisku-balioa (Value at Risk, VaR)

- VaR p probabilitate jakin bati dagokion pertzentila da finean. Adibidez, $VaR(0.05) = 2$ izanik, inbertsioan %0.05eko probabilitatea dago mozkinak 2 edo txikiagoak izateko. Hau da, 20 egunetatik batean espero da batean irabaziak 2 baino txikiagoak izatea.
- Arrisku-balioa (nazioartean ezagunagoa *Value at Risk, VaR* izenekin) inbertsioen arriskua neurtzeko adierazle bat da, eskuratuko den etekinari buruzkoa. Probabilitate txiki horiekin inbertsioaren alde txarrari, arriskuari alegia, erreparatzen dio.
- Eta beraz, VaR zenbat eta handiagoa, inbertsioa orduan eta hobea dela irizten da (izan ere, zerbait txarra esan behar badidate, izan dadila ahalik eta onena, kasu honetan ahalik eta handiena, errendimenduei buruz ari garelako).