Berretura-legeko banaketak

Josemari Sarasola

Gizapedia

Berretura-legea

X eta Y aldagaiak potentzia-lege batez loturik daudela esaten da, bata bestearen berretura batekin proportzionalki erlazionaturik dagoenean:

$$Y = KX^{\alpha}$$

Logaritmoak hartzen badira, erlazio hori lineal bihurtzen da:

$$\ln Y = \ln K + \alpha \ln X$$

Berretura-legeak estatistikan

Estatistikan, berretura-legeek x zorizko aldagaia eta horren $p(x)$ probabilitatea lotzen dituzte, berretzailea beti negatiboa izanik:

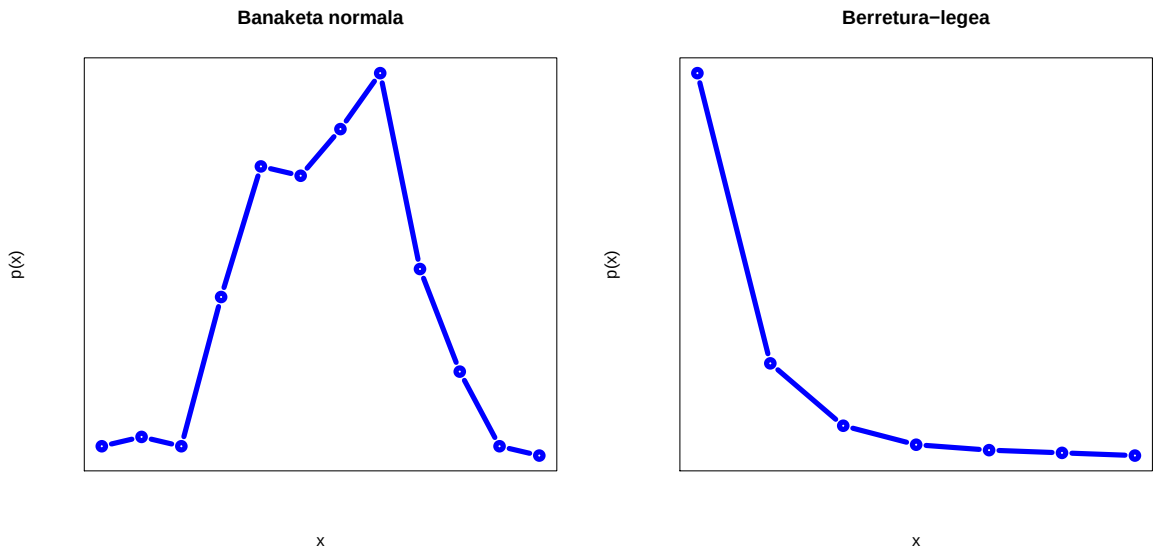
$$p(x) = Kx^{-\alpha}$$

Logaritmoak hartzen badira,

$$\ln p(x) = \ln K - \alpha \ln x$$

Beraz, α parametroak x balio baten probabilitatearen bilakaera adierazten du, x balioen arabera: α zenbat eta handiagoa, orduan eta azkarrago egiten du behera probabilitatea, hots, orduan eta ugariagoak dira balio txikiak, eta orduan eta urriagoak balio handiak.

Berretura-legeak estatistikan



Berretura legeen arabera, aldagaian balio txikia hartzen duten elementu asko daude, eta balio handia hartzen duten elementu gutxi.

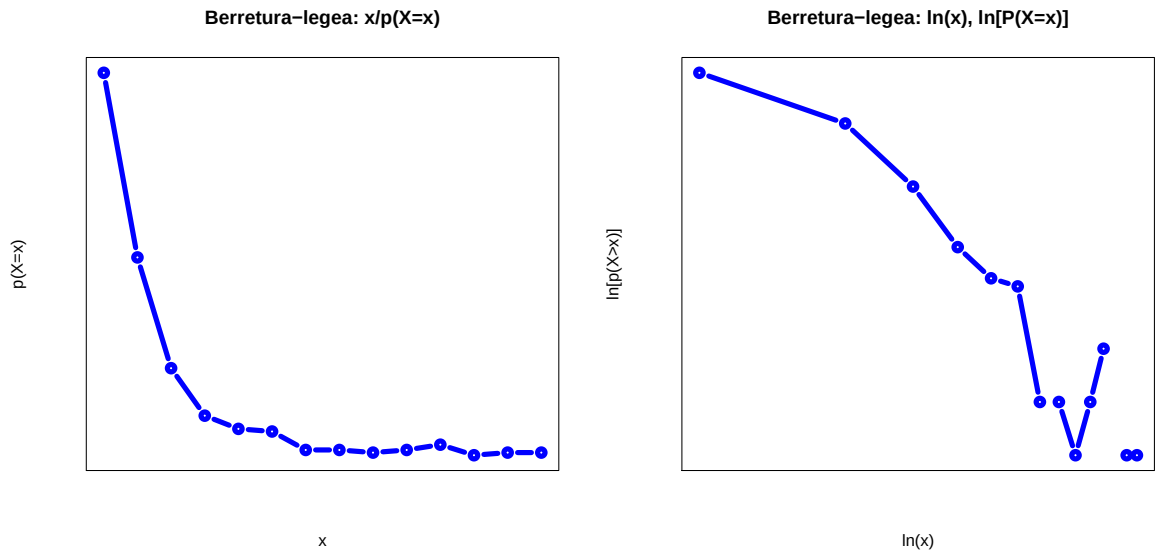
Berretura-legeen adibideak

Berretura-legeak tamaina eta kopuruekin loturiko aldagaietan agertu ohi dira:

- familien errentak: soldata apaleko familia asko, aberats gutxi batzuk;
- enpresetako langile kopuruak: enpresa asko langile gutxirekin, gutxi batzuk langile askorekin;
- liburu batetik saldutako aleak: salmenta handiko liburu gutxi, salmenta gutxiko liburu asko.

Berretura-legeak nola aurkitu

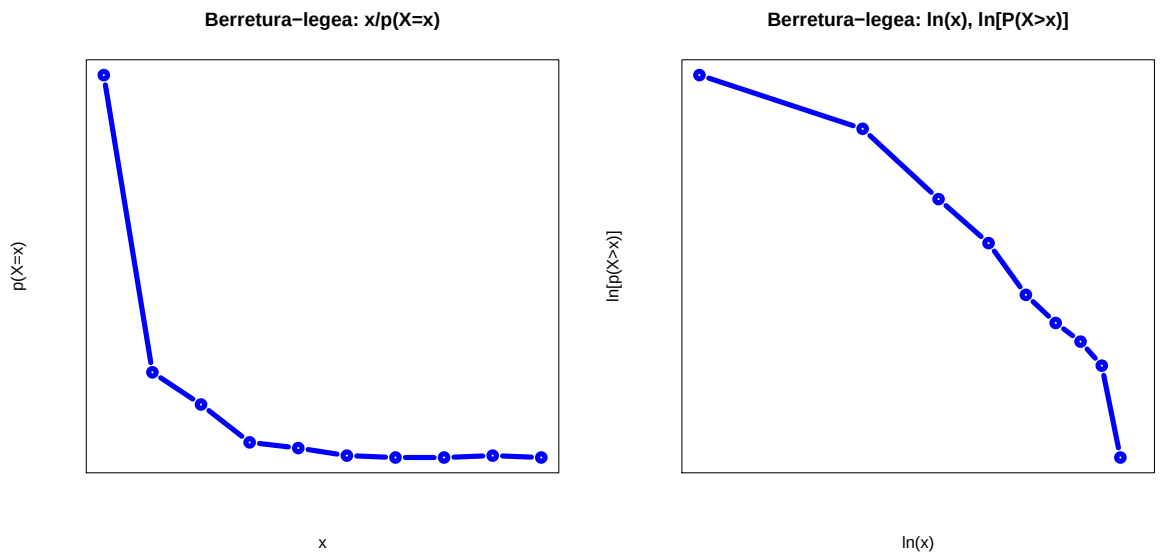
Banaketa normalaren ondoren, berretura-legeak gehien aurkitzen diren datu-egiturak dira.
Banaketa normala identifikatzeko, aski da kanpai itxura eta gutxi gorabeherako simetria topatzea. Baina, nola identifikatu berretura-legeak?
x eta $p(x)$ irudikatu ordez, horien logaritmoak irudikatzen dira.
Berretura-lege baten araberako erlazioa kontsideratzeko, gutxi gorabeherako zuzena suertatu behar da:



Berretura-legeak nola aurkitu

Azken grafikoan lerro zuzena ote dugu? Gauzak garbiago ikusteko,

- datu jarraituetarako histograma egin denean, zabalera konstanteko tarteak hartuz ordez, bilakaera logaritmikoa duten tarte-zabalerak eratzea (adibidez: 10-100-1000 edota 2-4-8-16)
- edo, egokiagoa izaten dena, $\ln x$ eta $\ln p(X > x)$ irudikatzea, hau da, gorako probabilitatearen ordez, hortik gorako probabilitateen logaritmoak irudikatzea. Horrela, garbiago ikusiko dugu lerro zuzenaren antzeko egitura berretura-legea dagoenean.



Berretura-legeen minimoa

- Hala eta guztiz ere, berretura-legeak normalean ez dira betetzen aldagaiaren eremu osoan.
- Adibidez, familien errenten kasuan, errenta minimo batetik behera irabazten dutenak oso gutxi dira eta ez asko; herrien biztanleria minimo bat ere egoten da (zaila da 10 edo bizilagun gutxiagoko herriak aurkitzea).
- Hori dela eta, berretura-legeak x_{min} balio minimo batetik aurrera aztertu eta aplikatzen dira.
- Horregatik, aldagaiak baino, aldagaietako isatsak edo muturrak dira berretura-legeen araberakoak (ohikoa da *power-law tail* esatea).

Dentsitate-funtzioa

Eremu jarraituan, froga daiteke berretura-legeen araberako dentsitate-funtzioak era honetakoak direla, betiere minimo batetik abiatzen garela suposaturik:

$$f(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha} \; ; \; x > x_{min} \; ; \alpha > 1$$

Ohartarazi behar da α parametroa 1 baino handiagoa izan behar dela derrigorrez, dentsitate-funtzioa behar bezala eratzeko. Hori ez da oztopoa horren aplikazio praktikorako, errealitatean oso arraroak direlako $\alpha < 1$ duten fenomenoak.

Banaketa-funtzioa

Probabilitateak errazago kalkulatzen dira formula hauekin:

$$F(x) = P[X < x] = 1 - \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

$$\overline{F}(x) = P[X > x] = \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

Itxaropena

$$\forall \alpha > 2, E[X] = \frac{\alpha - 1}{\alpha - 2} x_{min}$$

$\alpha \leq 2$ balioetarako, berriz, itxaropena ez da existitzen. Zehatzago, infinitua da, eta beraz dibergentea. Zer esan nahi du horrek? Probabilitateekin kalkulatu ordez, batezbestekoa datuekin kalkulatuko bagenu, lagin tamaina zenbat eta handiagoa, orduan eta batezbesteko aritmetiko handiagoa aterako litzatekeela, mugarik gabe n handitu ahala, infinituraino heldu arte.

Maximoaren itxaropena

n lagin-tamaina izanda,

$$E[X_{max}] = n\frac{1}{\alpha - 1}$$

Horrela, $\alpha > 1$ betetzen denez, maximoaren itxaropena beti goraka doa, lagin-tamaina handitu ahala. Baina, $\alpha \leq 2$ denean, maximoaren itxaropena era *leherkor* batean, esango genuke, egiten du goraka, formulan horrelako balioak ordeztuz ikus daitekeenez. Hortik, itxaropena finitua izateko $\alpha > 2$ izatearen beharra.

Kontzentrazioa: Lorenz kurba

Errentaren banaketari dagokionean esaterako, aberastasun osoaren zein A zati du biztanleriaren P portzentaje aberatsenak?

$$\forall \alpha > 2, A = P^{\frac{\alpha-2}{\alpha-1}}$$

Adibidez, %20 aberatsenen portzentajea hartzen bada:

α	A
2.1	0.86
2.3	0.69
2.5	0.58
2.7	0.51
2.9	0.46

Beraz, zenbat eta α txikiagoa, kontzentrazioa orduan eta handiagoa da.

Baldintzapeko banaketak

Berretura-lege baten baldintzapeko banaketa, $X > x_0$ delako baldintzapean, berretura-lege bat da, α parametro berdinarekin, eta x_{min} ordez, x_0 harturik:

$$P[X > x/X > x_0] = \frac{P[X > x]}{P[X > x_0]} = \frac{\left(\frac{x}{x_{min}}\right)^{1-\alpha}}{\left(\frac{x_0}{x_{min}}\right)^{1-\alpha}} = \left(\frac{x}{x_0}\right)^{1-\alpha}$$

Aurrekoa x_0 minimotzat duen berretura-lege baten banaketa-funtzioaren osagarria besterik ez da.

Pareto banaketa

Pareto banaketa berretura-legeen aukerako parametrizazioa besterik ez da, bereziki errentaren banaketarako erabiltzen dena. Izena banaketa hori aurkitu zuen Vilfreto Pareto (1848-923) italiar ekonomialariaren omenez darama.
Honela adierazten da bere banaketa-funtzioa:

$$F(x) = 1 - \left(\frac{x_{min}}{x}\right)^{\alpha}$$

Beraz, Pareto banaketaren α parametroa berretura-legearen banaketaren $\alpha - 1$ besterik ez da, kalkulu algebraiko sinple batez froga daitekeenez.
Idaztean, garrantzitsua da aipatzea erabiltzen ari garen parametrizazioa zehaztea.

Alfa parametroaren estimazioa

$$\hat{\alpha} = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]$$

Berretura-legeak eremu diskretuan: Zipf legea

Arestian ikasi dugun berretura legea zorizko aldagaia jarraituetarako erabiltzen da da. Beste kasu batzuetan aldagaia kualitatiboa edo diskretua da, hala nola hiri nagusienen tamaina, testu bateko hitzen azalpena, Gizapediako artikulu bisitak, 2 seme alaba edo gehiagoko familien kopurua..., baina berretura-legeko fenomeno berbera aurki dezakegu haietan:

- nazio bateko hiri handienak biztanle handia du, hortik aurrera hiriek gero eta biztanleria txikiagoa.
- hitz batzuk maiz errepikatzen dira, gehienak gutxitan;
- artikulu asko bisitatzen dira, gehienak gutxitan;
- 2 seme alabekin familia asko, hortik aurrera gero eta gutxiago.

Berretura-legeak eremu diskretuan: Zipf-en legea

George Kingsley Zipfek (1902-1950), Harvard Unibertsitateko hizkuntzalariak, ingelesezko testuen hitzen maiztasuna aztertu eta gutxigorabehera lege honi jarraitzen diotela aurkitu zuen:

$$f = \frac{0.1}{k}$$

k hitzaren heina da (1go hitz sarriena, 2gna, ...) eta f hitzaren maiztasuna. Horrela hitz sarriena ("the") 0.1/1=0.1=%10etan azalduko lizateke, bigarren sarriena 0.1/2=0.05=%5etan, ... Zipf-en legeak beti lotzen du heina maiztasun edo tamaina erlatiboarekin.

Berretura-legeak eremu diskretuan: Zipf-en legea

Zipf-en legearen adierazpen zehatza honelakoa da:

$$f = \frac{1}{k^a}$$

non f elementu edo balio bakoitzaren maiztasun edo tamaina erlatiboa den, beste elementuei buruz, k heina, eta a parametro bat, gehienetan 1 baino handiagoa izaten dena.
Elementu kopurua finkoa denean, maiztasun erlatiboen edo probabilitateen batura 1 izan dadin, aurreko maiztasunak edo tamainak N elementu guztien maiztasunen baturekin zatitzen da:

$$f_z = \frac{\frac{1}{k^a}}{\sum_{k=1}^{N} \frac{1}{k^a}}$$

Berretura-legeak eremu diskretuan: Zipf-en legea

Hala ere, guk ariketetan hasierako formulazio sinplea erabiliko dugu:

$$f = \frac{C}{k}$$

$$f_z = \frac{\frac{C}{k}}{\sum_{k=1}^{N} \frac{C}{k}}$$

Berretura-legeak eremu diskretuan: Zipf-en legea

Nola hatzematen diren Zipf-en legeak

k heinen eta f tamaina edo maiztasunen logaritmoak irudikatzen dira kartsiar diagrama batean. Lerro zuzen baten araberakoak baidra gutxi gorabehera, Zipf-en legea aplikatu daiteke.

Berretura-legeak eremu diskretuan: Zipf-en legea

Adibidez, 6 udalerriko eskualde batean, haietako biztanleriak modelizatzeko $f = \frac{3}{k}$ legea erabiliko balitz,

Heina	Tamaina erlatiboa (f)	Maiztasun erlatiboa (f_z)
1	3	0.40
2	1.5	0.20
3	1	0.13
4	0.75	0.10
5	0.6	0.08
6	0.5	0.07
	7.35	1

Hau da, herri populatuenean biztanleria osoaren %40 biziko litzateke.

Parametroaren estimazioa

Parametroa estimatzeko oso metodo sinplea erabiliko dugu, batere zorrotza ez den arren: Zipf legeko f maiztasuna maiztasun enpirikoarekin berdinduko dugu, edozein hein harturik. Aurreko adibidean, bigarren herri populatuenean biztanleriaren %25 biziko balitz:

$$\frac{C}{2} = 0.25 \rightarrow \hat{C} = 0.5$$

Aukeratutako heinaren arabera, estimazio ezberdinak izango dira noski. Estimazio onena maiztasun teoriko eta enpirikoen arteko diferentzia karratuen batura erabil liteke (ikus ariketak).