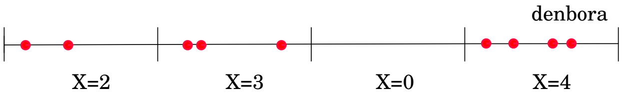
Poisson prozesuak eta loturiko banaketak

Josemari Sarasola

Gizapedia

Poisson banaketa

Poisson banaketak epe batean (minutu batean, ordu batean, egun batean) gertaera puntualen kopuru bat (matxura kopurua, istripu kopurua, igarotzen den ibilgailu kopurua, webgune bateko bisita kopurua) izateko probabilitatea ematen du.



Baldintza hauek bete behar dira Poisson prozesu baterako:

- gertaera puntualak zoriz eta independentziaz gertatu behar dira edozein unetan;
- gertaera puntualak λ tasa konstante baten arabera gertatu behar dira.

Poisson banaketa

Poisson banaketak adierazten du denborako epe batean zenbat gertaera gertatzen diren. Epe bakoitzeko gertaera kopurua honela banatzen da:

Probabilitate-funtzioa

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!};\ x = 0, 1, 2, \ldots (\infty - raino)$$

Adierazpena, itxaropena eta bariantza

$$X \sim P(\lambda) \begin{cases} \mu = \lambda \\ \sigma^2 = \lambda \end{cases}$$

Parametro bakarra du: λ.

Poisson banaketa

R softwareaz

$$X \sim P(2.2)$$

- `dpois(3,2.2) #P[X=3]`
- `ppois(5,2.2) #P[X<=5]`
- `ppois(4,2.2,lower.tail=FALSE] #P[X>4]`
- `1-ppois(4,2.2) #P[X>4]`
- `x=0:10`
- `dpois(x,10,0.2) #prob sinple batzuk (denak ez, infinitu baitira)`

Poisson banaketa binomialaren limite moduan

Banakuntza binomiala Poisson banakuntzaren bitartez hurbildu daiteke, n handia eta p txikia denean:

$$B(n,p) \to P(\lambda = np)$$

Hurbilketa ona da $n \geq 20$ eta $p \leq 0.05$ direnean, eta oso ona $n \geq 100$ eta $np \leq 10$ kasuetan.

- 6 magnitudeko lurrikara 1000 urtetan behin gertatzen da Iberiar Penintsulan. Magnitude horretako lurrikararen itzuli-aldia 1000 urtekoa da, orduan.
- Hildakoa dakarren auto-istripua 15 egunetan behin gertatzen da Gipuzkoan. Istripu horien itzuli-aldia 15 egun da, orduan.

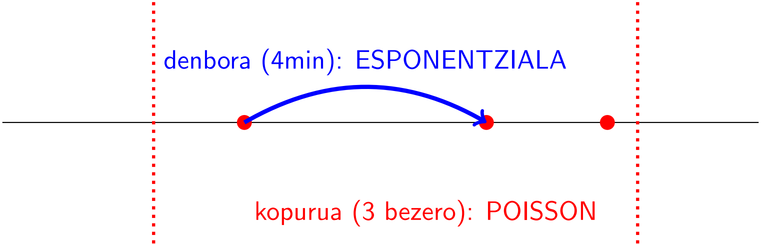
Itzuli-aldia (ia)

$\lambda = 1$ izateko behar den denbora da. Beraz, t denborako lambda λ_t izanik, $ia = \frac{t}{\lambda_t}$.

Adibidez, 20 hilabetero 4 matxura gertatzen badira batezbeste, matxuren itzuli-aldia 20/4=5 hilekoa da.

Banaketa esponentziala

Poisson prozesu batean gertaera kopurua zorizkoa denez, gertaera batetik bestera igarotzen ere denbora ere zorizko aldagaia da. Gertaera kopurua Poisson banaketaren arabera banatzen da. Gertaera batetik besterako denbora banaketa esponentzialaren arabera banatzen dela esango dugu:



Banaketa esponentziala

Ondoz ondoko gertaeren arteko denbora azaltzen duen banaketa esponentziala zehazten dugu:

Dentsitate eta banaketa funtzioak

$$f(x) = \lambda e^{-\lambda x};\ \ x > 0$$

$$F(x) = P(X < x) = 1 - e^{-\lambda x};\ \ x > 0$$

Itxaropena eta bariantza

$$X \sim Exp(\lambda):\ \ \mu = \frac{1}{\lambda};\ \sigma^2 = \frac{1}{\lambda^2}$$

Oharra: denbora jarraitua denez: $P[X < x] = P[X \leq x]$

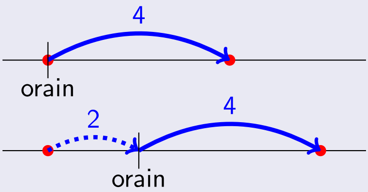
Banaketa esponentziala

Itxaropenaren emaitza guztiz logikoa da: $\lambda = 4$ bada orduko, adibidez; batetik bestera batezbestez $1/\lambda = 0.25\ ordu$ izango da. Ohartu kointzidentziaz: lambda orduko bada, esponentzialaren emaitzak ere ordutan izango dira.
Hain zuzen, Poissonen banaketako λ parametroaren aldia (minutua, ordua, ...) izango da banaketa esponentzialean x baliorako denbora unitatea. Adibidez, λ minutuko bada, x denbora minututan kontatu behar da esponentzialean.

Oroimen eza

Gertaerak zoriz eta independentziaz suertatzen direnez, gertaera batetik bestera 4 minutu baino gehiago izateko probabilitatea eta 2 minutu igaro direlarik beste 4 minutu gutxienez (beraz, guztira 6 minutu baino gehiago) izateko probabilitatea berdinak dira:

$$P[X > 4] = P[X > 6/X > 2]$$



Hau da, 2 minutu igarotzeak ez du informaziorik ematen hurrengo gertaera izan arte igaroko den denborari buruz.

Banaketa esponentziala

Oroimen-eza

Orokorrean, d denbora pasata, x denbora gehiago (guztira, beraz, $d+x$) pasatzeko probabilitatea, eta hasieratik x denbora pasatzeko probabilitatea berdinak dira:

$$P(D > d + x / D > d) = P(D > x)$$

Beraz, banaketa esponentzialak gertaera batetik besterako denbora azaltzen du, baina baita ere, orokorrago, hurrengo gertaera izan arteko denbora.

Banaketa esponentziala

Iraupenak modelizatzeko eredu gisa

Gertaeren arteko denborarako ez ezik, banaketa esponentziala gailu eta makinen iraupen eta bizitza erabilgarriari buruzko probabilitateak kalkulatzeko ere erabiltzen da. Aldagai horietarako erabiltzen denean, *oroimen-ezak desgaste edo neke eza inplikatzen du*, aurretik iragandako denborak ez baitu aldatzen ondorengo denbora eta gertaeren probabilitateetan.

Banaketa esponentziala

Iraupenak modelizatzeko eredu gisa

Hala, auto berri batek 2 urte irauteko probabilitatea eta 10 urteko auto batek beste bi urte irauteko probabilitatea berdinak lirateke. Hau da, hondatze-tasa konstantea da denbora zehar banakuntza esponentzialarekin.
Errealitatean gauzak ez dira hala izaten, eta *hondatze-tasa* goraka doa denbora igaro ahala (10 urteko autoak hondatzeko aukera gehiago lituzke berriak baino). Kasu horietarako, iraupenak modu errealistago eta malguago batez irudikatzen duen eredua Weibull banaketa da.

Iraupenak modelizatzeko eredu gisa

Beste kasu batzuetan, hondatze-tasa beherakorra da hasiera batean, egonkorra gero eta gorakorra amaieran. Adibidez, pertsonak jaioberritan gaixotzeko arriskua handia dute, baina arrisku hau beheraka doa haurra hazten doanean (tasa beherakorra), heltzean hondatze-tasa nahiko egonkorra da (aldi horretarako soilik banakuntza esponentziala egokia litzateke) eta zahartzaroan, goraka egiten du urteekin batera (hondatze-tasa gorakorra). Egoera horiek guztiak barnehartzen dituen ereduetako bat Gompertz banaketa da.

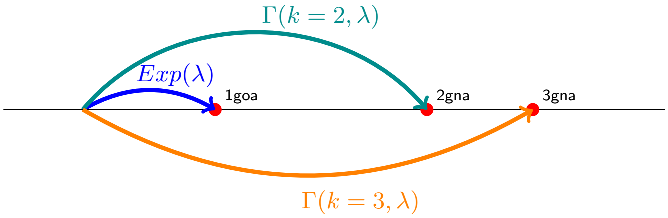
Gamma banaketa

Poisson prozesu batean gertaera kopurua zorizkoa denez, gertaera batetik k-garren gertaera izan arte igarotzen ere denbora ere zorizko aldagaia da (gogoratu *hurrengo* gertaera arteko denbora $Exp(\lambda)$ dela).

k-garren gertaera arteko denboraren probabilitate banaketa Gamma banaketa da, bi parametrorekin:

- λ, betiko esanahiarekin
- k, hurrena gertatuko den k-garren gertaera.

Labur (Γ, gamma letra maiuskula): $X \sim \Gamma(k, \lambda)$

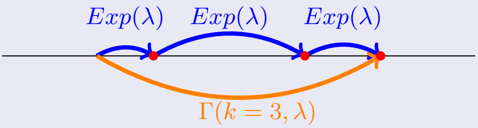


Gamma banaketa

Lotura banaketa esponentzialarekin

$k = 1$ kasuan banaketa esponentzialaren kasuan gaude (lehen gertaera hurrengoa baita): $\Gamma(k = 1, \lambda) \equiv Exp(\lambda)$.

Gainera, $\Gamma(k, \lambda)$ Gamma banaketa k banaketa esponentzial ($Exp(\lambda)$) independenteen batura da:



$$Exp(\lambda) + Exp(\lambda) + Exp(\lambda) = \Gamma(k = 3, \lambda)$$

Batezbestekoa eta bariantza $\Gamma(k, \lambda)$ banaketan

Beraz $\Gamma(k, \lambda)$ banaketaren itxaropena eta bariantza k esponentzialen itxaropenen ($1/\lambda$) eta bariantzen ($1/\lambda^2$) batura da:

$$X \sim \Gamma(k, \lambda) : \ \mu = \frac{k}{\lambda}; \ \sigma^2 = \frac{k}{\lambda^2}$$

Dentsitate edo banaketa funtzioarekin probabilitateak kalkulatzea oso konplexua da. Ikusi zein izugarri den Gamma banaketaren dentsitate funtzioa:

$$f(x;k,\lambda) = \frac{x^{k-1}e^{-\lambda x}}{\frac{1}{\lambda^k}\Gamma(k)} \quad x > 0; k, \lambda > 0$$

non $\Gamma(k)$ funtzio matematiko berezi bat den.

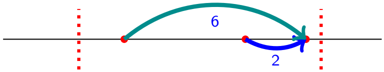
Askoz ere aiseago, probabilitateak gertaera kopuruaren terminoetan pentsatuz kalkulatzen dira, hots Poisson banaketaren bitartez.

Adibidez, $X \sim \Gamma(k = 3, \lambda = 0.3)$ izanik, nola kalkulatu $P[X > 10]$?
3garren gertaerara bitartean 10 denbora-unitate baino gehiago izateko, 10 denbora-unitateko epean 3 gertaera baino gutxiago izan behar dira, hots, 2 edo gutxiago. 10 denbora-unitatean, $\lambda = 0.3 \times 10 = 3$ dugunez,

$$P[X > 10] = \frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!}$$

Poisson prozesuak: lasterbidea

Poisson prozesu bat (adibidea):



Poisson prozesu bati atxikitako banaketak (λ: batez besteko gertaera kop., k: hurrengo *k* gertaera):

Izena	Notazioa	Aldagaia	Adibidea
Poisson	$P(\lambda)$	Gertaera kopurua aldi batean	x=3
Exponentziala	$G(p)$	Denbora hurrengo gertaerarako	x=2
Gamma	$\Gamma(k,\lambda)$	Denbora hurrengo k-gn gertaerarako	k=2,x=6

Izena	Formula	Euskarria
$P(\lambda)$	$P[X=x]=\frac{e^{-\lambda}\lambda^x}{x!}$	$x=0,1,2,..,\infty$
$Exp(\lambda)$	$P[X>x]=1-e^{-\lambda x}$	$x>0$
$\Gamma(k,\lambda)$	konplexua (Poissonekin ebazten da)	$x>0$