Banaketa normala eta limitearen teorema zentrala

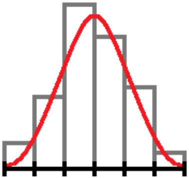
Josemari Sarasola

Estatistika enpresara aplikatua

Banaketa normala

Estatistikan gehien erabiltzen den banaketa da arrazoi hauengatik:

- kanpai itxurakoa eta simetrikoa da, eta horrela errealitatean maiz agertzen diren fenomeno askotara aplika daiteke (hortik datorkio *normal* izena);



- beste hainbat probabilitate-banaketen limitea da;
- eta propietate matematiko interesgarriak ditu.

Banaketa normala

XVIII. mendetik ezagutzen da, Abraham de Moivre eta Pierre-Simon de Laplace matematikariei esker. Carl Friedrich Gauss matematikariak aplikatu zuen lehen aldiz, errore astronomikoen arloan. Horregatik, Gauss-en ezkila ere deitzen zaio.

Irudia: Alemaniako markoaren billete bat, non Gauss eta banaketa normala agertzen diren.

Banaketa normala

Definizioa, notazioa, parametroak, simetria

- Dentsitate-funtzioa (ez ikasi, praktikan ez baita erabiltzen):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \; -\infty < x < \infty$$

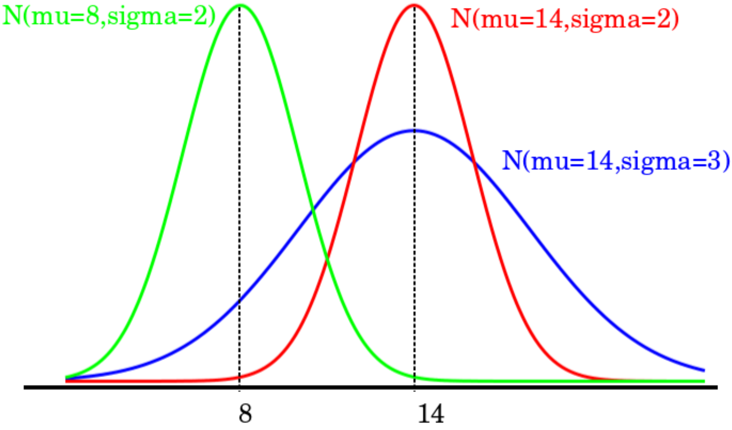
- Parametroak:

$$X \sim N(\mu, \sigma)$$

  μ : itxaropena σ : desbideratze estandarra

- μ itxaropenaren inguruan simetrikoa da (erabilgarria probabilitateak kalkulatzeko)

Banaketa normala



Banaketa normala

Banaketa normal estandarra

$Z \sim N(\mu = 0, \sigma = 1)$ banaketa normal *estandarra* da. Beti Z deitzen zaio. Oso garrantzitsua da, beste banaketa normalei buruzko probabilitateak kalkulatzeko patroi edo estandar gisa erabiltzen delako, estandarketa izeneko eragiketaz.

Estandarketa

Probabilitateak kalkulatzeko, banaketa normal guztiak *estandar* bilakatu behar dira, *estandarketa* izeneko prozesuaz:

$$X \sim N(\mu, \sigma) \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Hots, banaketa normal bati batezbestekoa kendu eta zati desbideratzea egiten bazaio, banaketa normal estandar bilakatzen da.

Propietatea: aldakuntza lineala

$X \sim N(\mu, \sigma)$ banaketa izanik, $Y = a + bX$ zorizko aldagaia ere banaketa normalaren araberakoa izango da, parametro hauekin:

$$Y \sim N(a + b\mu, |b|\sigma)$$

Propietatea: ugalkortasuna

- $X_1 \sim N(\mu_1, \sigma_1)$
- $X_2 \sim N(\mu_2, \sigma_2)$
- ...
- $X_n \sim N(\mu_n, \sigma_n)$ banaketak izanik, elkarrekiko independenteak, honako hau betetzen da;

$$Y = X_1 + X_2 + \ldots + X_n \sim N\left(\mu_1 + \mu_2 + \ldots + \mu_n, \sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}\right)$$

Hau da, banakuntza normalen batura banaketa normalaren araberakoa da, batezbestekoa batezbestekoen batura eta desbideratzea *bariantzen baturaren erroa* izanik.

Ohartu: $\sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2} \neq \sigma_1 + \sigma_2 + \ldots + \sigma_n$ (bariantzak gehitzen dira beti!)

R softwareaz

```
pnorm(1.45) #P[Z<1.45]
1-pnorm(1.45) #P[X>1.45]
pnorm(1.45,lower.tail=FALSE] #P[X>1.45]
pnorm(4,mean=3,sd=1.5) #P[X>4], X:N(3,1.5)
qnorm(0.90) #P[Z<z]=0.9,z?
qnorm(0.90,mean=3,sd=1.5) #P[X<x]=0.9, x?,X:
N(3,1.5)
```

De Moivre-Laplace teorema: binomialaren hurbilketa normala

Banaketa binomial batean, n handia bada $(n \geq 30)$ p txikia izan gabe $(np \geq 10)$ -izan ere, beste kasuetan, p txikiarekin, binomialaren hurbilketa Poisson bidez egiten baita, $\lambda = np$ harturik-, banaketa normala erabil daiteke probabilitate binomialak hurbiltzeko:

$$B(n,p) \xrightarrow[n \geq 30]{} N(\mu = np, \sigma = \sqrt{npq})$$

Hurbilketa honi De Moivre-Laplace teorema deitzen zaio.

Banaketa normala

Poisson banaketaren hurbilketa normala

Poisson banaketa batean, λ handia bada $(\lambda \geq 30)$, hurbilketa gisa banaketa normala erabil daiteke:

$$P(\lambda) \xrightarrow[\lambda \geq 30]{} N(\mu = \lambda, \sigma = \sqrt{\lambda})$$

Jarraitutasun-zuzenketa

De Moivre-Laplace teoreman nahiz Poisson banaketaren hurbilketa normalean, finean egiten duguna da banaketa diskretu bat banaketa jarraitu baten bitartez hurbildu. Hurbilketa hori zehatzagoa izan dadin jarraitutasun-zuzenketa (egin klik estekan gehiago jakiteko) erabiltzen da. Zuzenketa nola burutzen den ikasteko, hona adibide batzuk:

- $P[X = 10] = P[9.5 < X < 10.5]$
- $P[X \leq 10] = P[X < 10.5]$
- $P[X \geq 10] = P[X > 9.5]$
- $P[X > 10] = P[X > 10.5]$
- $P[X < 10] = P[X < 9.5]$

Zuzenketa ez egiteagatik sortzen den errorea oso txikia da, eta horregatik beharrezkoa izango da soilik zehaztasun handia behar denean kalkuluetan.

Limitearen teorema zentrala (LTZ)

Badakigu banaketa normalen batura banaketa normalari jarraiki banatzen dela, batzen diren banaketak elkarrekiko independenteak badira (ugalkortasunaren propietatearengatik). Eta zer batzen diren banaketak normalak ez badira (uniformeak, esponentzialak, bestelakoak badira)?

Normalak ez diren banaketen batura normal banatzen baita ere, baina bi baldintzekin:

1. batugai diren banaketak independenteak izan behar dira beraien artean, normalen baturan bezalaxe, eta
2. (hau berria da) gehitzen diren banaketen kopurua aski handia izan beha da, 30 edo handiagoa orokorrean.

Emaitza horri Limitearen Teorema Zentrala deitzen zaio. De Moivre-Laplace teorema eta Poissonen hurbilketa normala LTZaren kasu bereziak besterik ez dira.

Limitearen teorema zentrala (LTZ)

- $X_1 \sim ?(\mu_1, \sigma_1)$
- $X_2 \sim ?(\mu_2, \sigma_2)$
- ...
- $X_n \sim ?(\mu_n, \sigma_n)$ harturik, edonolakoak baina batezbesteko eta bariantza jakinekin, elkarrekiko independenteak, honako hau betetzen da, n kopurua aski handia bada $(n \geq 30)$;

$$Y = X_1 + X_2 + \ldots + X_n \sim N\left(\mu_1 + \mu_2 + \ldots + \mu_n, \sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}\right)$$

Hau da, banaketa askoren batura, batezbesteko eta bariantza jakinekin, eta elkarrekiko independenteak izanik, normal banatzen da.