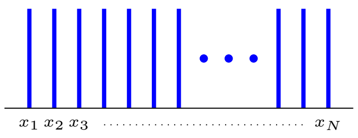
Banaketa uniformea

Josemari Sarasola

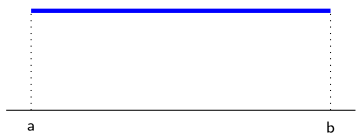
Gizapedia

Banaketa uniformea

Banaketa uniformea balio posible guztiei probabilitate berdina ematen dien probabilitate-banaketa da. Bi erako banaketa uniformeak bereizten dira:



Banaketa uniforme diskretua



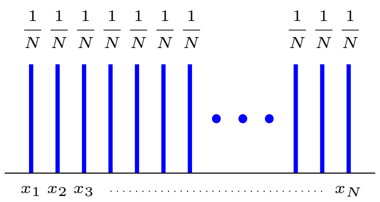
Banaketa uniforme jarraitua

Banaketa uniformeak aldagai bati buruz erabateko ziurgabetasuna dagoenean aplikatzen dira, erabateko ezjakintasunean balio posible guztiei probabilitate berdina esleitu behar zaielako. Populazio batetik elementuak zoriz aukeratzean ere erabiltzen dira, zorizko elementu guztiek probabilitate berdina dutelako. Azkenik, zorizko zenbakiak sortzeko ere erabiltzen dira, zorizko zenbakiak definizioz probabilitate berdina duten horiek direnez.

Banaketa uniforme diskretua

Probabilitate-funtzioa

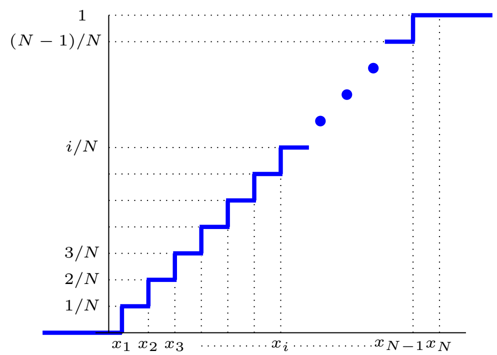
$$P[X=x]=\frac{1}{N}\ ;\ x=x_1,x_2,\cdots,x_N$$



Banaketa uniforme diskretua

Banaketa-funtzioa

$$F(x)=P[X\leq x_i]=\frac{i}{N}\ ;\ x_i=x_1,x_2,\cdots,x_N$$



Banaketa uniforme diskretua

Adierazpena, itxaropena eta bariantza

$$X \sim U(x_1, x_2, \ldots, x_N) \begin{cases} \mu = \dfrac{x_1 + x_N}{2} \\ \sigma^2 = \dfrac{(x_N - x_1 + 2)(x_N - x_1)}{12} \\ = \dfrac{(x_N - x_1 + 1)^2 - 1}{12} \end{cases}$$

Banaketa uniforme diskretua

Maximoaren banaketa

Banaketa uniforme diskretu bat n aldiz gauzatzen da. Nola banatzen da n balio horietatik handiena?

- n balioetatik handiena x_i baino txikiagoa izango da haiek guztiak x_i baino txikiagoak direnean:

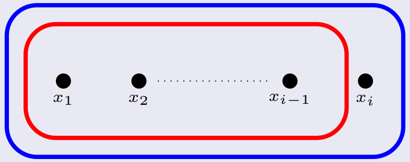
$$P[X_{max} \leq x_i] = \frac{i}{N} \times \frac{i}{N} \times \cdots \times \frac{i}{N} = \left(\frac{i}{N}\right)^n$$

- n balioetatik handiena x_{i-1} baino txikiagoa izango da haiek guztiak x_{i-1} baino txikiagoak direnean:

$$P[X_{max} \leq x_{i-1}] = \left(\frac{i-1}{N}\right)^n$$

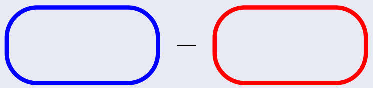
Banaketa uniforme diskretua

Maximoaren banaketa



Beraz,
n balioetatik handiena x_i izateko probabilitatea hau da:

$$P[X_{max} = x_i] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n$$



Banaketa uniforme diskretua

Maximoaren banaketa: adibidea

Dado bat 4 aldiz botata, eman zenbaki handiena 6,5,4,3,2 eta 1 izateko probabilitateak:

- $P[X_{max} = 6] = \left(\frac{6}{6}\right)^4 - \left(\frac{5}{6}\right)^4 = 0.5177$
- $P[X_{max} = 5] = \left(\frac{5}{6}\right)^4 - \left(\frac{4}{6}\right)^4 = 0.2847$
- $P[X_{max} = 4] = \left(\frac{4}{6}\right)^4 - \left(\frac{3}{6}\right)^4 = 0.1350$
- $P[X_{max} = 3] = \left(\frac{3}{6}\right)^4 - \left(\frac{2}{6}\right)^4 = 0.0501$
- $P[X_{max} = 2] = \left(\frac{2}{6}\right)^4 - \left(\frac{1}{6}\right)^4 = 0.0115$
- $P[X_{max} = 1] = \left(\frac{1}{6}\right)^4 - \left(\frac{0}{6}\right)^4 = 0.0007$
- Probabilitatea beheraka doa. Logikoa da.

Banaketa uniforme diskretua

Minimoaren banaketa

Era berean,
n balioetatik txikiena x_i izateko probabilitatea hau da:

$$P[X_{min} = x_i] = P[X_{min} \geq x_i] - P[X_{min} \geq x_{i+1}]$$
$$= \left(\frac{N-(i-1)}{N}\right)^n - \left(\frac{N-i}{N}\right)^n$$

Banaketa uniforme diskretua

Aplikazioak: laginketa populazio finituetan

- Populazio finitu batetik lagin bat aukeratzean, zorizko laginketa egiten bada, elementu guztiek aukeratuak izateko probabilitate berdina dute. Beraz, populazio horretarako eredu egokia banaketa uniforme diskretua da.
- Zorizko laginketa itzuleraz egiten bada, n tamainako lagin jakin bat ateratzeko probabilitatea hau da:

$$P[x_1, x_2, \ldots, x_n] = \frac{1}{N} \times \frac{1}{N} \times \ldots \times \frac{1}{N} = \left(\frac{1}{N}\right)^n$$

- Zorizko laginketa itzulerarik gabe egiten bada, n tamainako lagin bat ateratzeko probabilitatea hau da, banaketa hipergeometrikoa baliatuz:

$$P[x_1, x_2, \ldots, x_n] = \frac{1}{\binom{N}{n}} = \frac{1}{N} \times \frac{1}{N-1} \times \ldots \times \frac{1}{N-(n-1)} \times n!$$

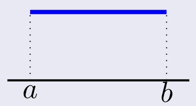
Banaketa uniforme diskretua

Aplikazioak: Alemaniako tankeen problema

- Estatistikako problema klasikoa da, II. Mundu Gerran planteatutakoa.
- Alemaniako tankeek matrikula bat daukate 1-etik N zenbaki ezezagun batera.
- Aliatuei interesatzen zaie N zenbaki hori ezagutzea. Horretarako, suntsitutako tankeen matrikulak jasotzen dituzte.
- Eredua banaketa uniforme diskretua da: $1, 2, \ldots, N$ tankeek probabilitate berdina dute suntsituak izateko.
- Banaketa uniforme horretatik n tanke jasota, matrikula maximoa nola banatzen den kalkulatzen da, N estimatzeke eta itzulerarik gabe, eta horren itxaropena kalkulatu:
  $E[X_{max}] = \dfrac{n(N+1)}{n+1}$
- Suertatu den maximoa, x_{max}, itxaropen horrekin berdindu eta N bakantzen da, tanke kopurua estimatzeko:
  $x_{max} = \dfrac{n(N+1)}{n+1} \rightarrow \hat{N} = x_{max} + \dfrac{x_{max}-n}{n}$
- Adibidez, suntsitutako tankeen matrikulak: 82,123,345,614.
- $\hat{N} = 614 + \dfrac{614-4}{4} = 766.5$ tanke dituztela estimatzen da.

Banaketa uniforme jarraitua

Dentsitate-funtzioa



$$f(x) = \frac{1}{b-a} \;;\; a < x < b$$

$$F(x) = P[X < x] = \frac{x-a}{b-a} \;;\; a \leq x \leq b$$

Adierazpena, itxaropena eta bariantza

$$X \sim U(a,b)$$

$$X \sim U(a,b) \begin{cases} \mu = \dfrac{a+b}{2} \\[2ex] \sigma^2 = \dfrac{(b-a)^2}{12} \end{cases}$$

- Itxaropenaren balioa guztiz intuitiboa da: tarteko balio guztiak probabilitate berekoak direnez, itxaropena erdi-erdian izango da.
- Tartea zenbat eta zabalagoa, sakabanatzea gero eta handiagoa, eta orduan bariantza ere handiagoa izango da.

Maximoaren banaketa

$U(a,b)$ banaketa uniforme jarraitu batetik n balio jakin harturik, haietatik M maximoa nola banatzen da?

- Banaketa-funtzioa:
$$F(M=x) = P[M<x] = \left(\frac{x-a}{b-a}\right)^n \;;\; a \le x \le b$$
- Itxaropena: $E[M] = a + \dfrac{n(b-a)}{n+1}$
- Adibidez, $U(0,10)$ banaketa batean batek maximoa estimatu nahiko balu 4 datuetatik lortutako maximoa harturik, 4 datu horiekin soilik batezbestez $0+\frac{4\times 10}{5}=8$ baliora emango luke, hau da, benetako maximoaren %80. 9 daturekin berriz, benetako balioaren %90era helduko litzateke.

Minimoaren banaketa

$U(a,b)$ banaketa uniforme jarraitu batetik n balio jakin harturik, haietatik m minimoa nola banatzen da?

- Banaketa-funtzioa:
$$F(m=x) = P[m<x] = 1 - \left(\frac{b-x}{b-a}\right)^n \;;\; a \le x \le b$$
- Itxaropena: $E[m] = a + \dfrac{b-a}{n+1}$
- Adibidez, $U(0,10)$ banaketa batean batek maximoa estimatu nahiko balu 4 datuetatik lortutako maximoa harturik, 4 datu horiekin soilik batezbestez $0+\frac{10}{5}=2$ baliora emango luke. 10 daturekin berriz, 1 emango luke batezbestez.

Ibiltartearen banaketa

$U(a,b)$ banaketa uniforme jarraitu batetik n balio jakin harturik, haietatik R ibiltartea nola banatzen da?

- Banaketa-funtzioa ($0 \le x \le (b-a)$ balioetarako):
$$F(R=x) = P[R<x] = n(\tfrac{x}{b-a})^{n-1}(\tfrac{(b-a)-x}{b-a}) + (\tfrac{x}{b-a})^n$$
- $E[R] = (b-a)\dfrac{n-1}{n+1}$
- Adibidez, $U(0,10)$ banaketa batean batek 10-0=10 ibiltartea estimatu nahiko balu 4 datuetatik lortutako ibiltartea harturik, 4 datu horiekin soilik batezbestez $0+10\frac{3}{5}=6$ baliora emango luke, benetako ibiltartearen %60. 10 daturekin berriz, 9/11=8.1 emango luke batezbestez.

Banaketa uniforme jarraitua

Banaketa uniforme estandarra

$$X \sim U(0,1)$$

Banaketa uniforme estandarra (0,1) tarteko *zorizko zenbakiak* sortzeko oinarri bezala erabiltzen da, kalkulagailuan SHIFT+RAN# sakatuz.

Simulazio estokastikoa

Simulazio estokastikoa banaketa batetik datuak *artifizialki* sortzea da, datu horiek errealak balira bezala (ikus, Gizapedian, Simulazio estokastikoa).
Banaketa uniforme jarraitu bat simulatzea oso erraza da beste banaketen aldean: zorizko zenbakiek $U(0,1)$ banaketa batetik datoz, eta $U(a,b)$ simulatzeko, aski da aldagai aldakuntza bat burutzea:
$U(a,b) = a + (b-a)U(0,1)$
Beraz, simulazioak sim izendatuz: $sim_{U(a,b)} = a + (b-a)sim_{U(0,1)}$