Statistika

- Donosi zaključke o fenomenima na osnovu uzoraka iz iskustva
  - Individualni slučajevi mogu odstupati od prosečnog (tipičnog)
  - Vršimo posmatranje velikog broja slučajeva da bismo otkrili zakonitost
  - Skup elemenata koji posmatramo se naziva *populacija*
  - Za svaki element populacije posmatramo određenu numeričku karakteristiku koju nazivamo *obeležjem*
  - Npr. celokupna proizvodnja fabrike sijalica čini jednu populaciju, a obeležje svake sijalice je „dužina života" izražena u časovima

- Osnovni zadatak statistike: *za datu populaciju naći distribuciju datog obeležja*
  - Npr. želeli bismo da znamo prosečan životni vek sijalice i koliko jedna individualna sijalica može da odstupa od tog proseka

Uzorak

- Broj elemenata populacije može da beskonačan. Merenje obeležja može da bude skupo, teško i vremenski zahtevno
- Zato se uzima samo jedan (konačan) deo populacije – *uzorak*
- Pretpostavka je da zaključci dobijeni na osnovu uzorka važe za celu populaciju
- Da bi ovo važilo, uzorak mora biti *reprezentativan*

Reprezentativan uzorak

- Uzorak je reprezentativan ako način selekcije elemenata ne zavisi od obeležja koje posmatramo
  1. Svaki element populacije mora da ima jednaku šansu da uđe u uzorak
  2. Uzorak mora da bude dovoljno brojan

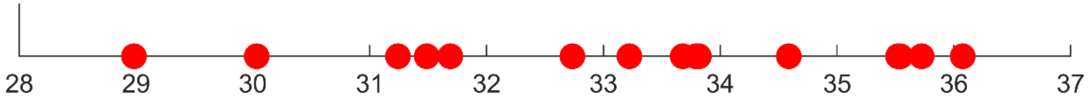
- Za tačke izvučene nezavisno jedna od druge iz iste distribucije se kažu da su *nezavisne i jednako distribuirane* (independent and identically distributed, **IID**)

- Primer: bacanje novčića. Prvo bacanje novčića ne utiče na ishod drugog bacanja

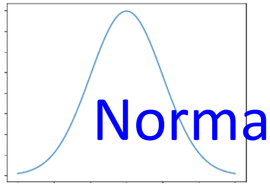
Estimacija parametara modela

- Donošenje zaključaka često se svodi na ocenu nekih nepoznatih parametara verovatnosne mere, na osnovu podataka

1. Imamo uzorak



2. Pretpostavimo da uzorak potiče iz neke distribucije $f(x_n|\theta)$ koja zavisi od nekih nepoznatih parametara θ



Normal $\mathcal{N}\left(x^{(n)}|\mu,\sigma^2\right)$

3. Naš zadatak jeste da odredimo θ

μ,σ

Metod maksimalne verodostojnosti

- Ideja: odabrati θ tako da verovatnoća realizacije dobijenog uzorka $x = \left(x^{(1)}, \ldots, x^{(N)}\right)$ bude najveća

- Pretpostavke:
  - Svaki element $x^{(n)}$ je odabran iz distribucije $f(x^{(n)}|\theta)$
  - Svaki element je odabran nezavisno jedan od drugog

- Zato, združena verovatnoća odabranog uzorka je:

$$L(\theta) = \prod_{n=1}^{N} f(x^{(n)}|\theta)$$

Metod maksimalne verodostojnosti

$$L(\theta) = \prod_{n=1}^{N} f(x^{(n)}|\theta)$$

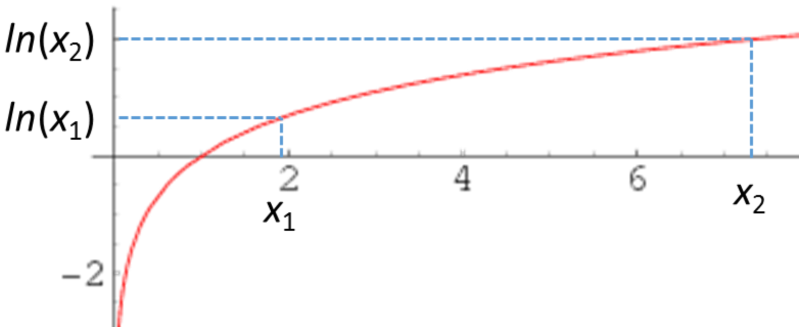
Funkcija verodostojnosti

Zavisi od θ

- Zadatak: pronaći vrednost θ koja je maksimizuje $L(\theta)$
- U maksimumu je $\frac{\partial}{\partial\theta} L(\theta) = 0$

Metod maksimalne verodostojnosti

- U praksi se minimizuje $-\log L(\theta)$
- Logaritam je monotono rastuća funkcija:

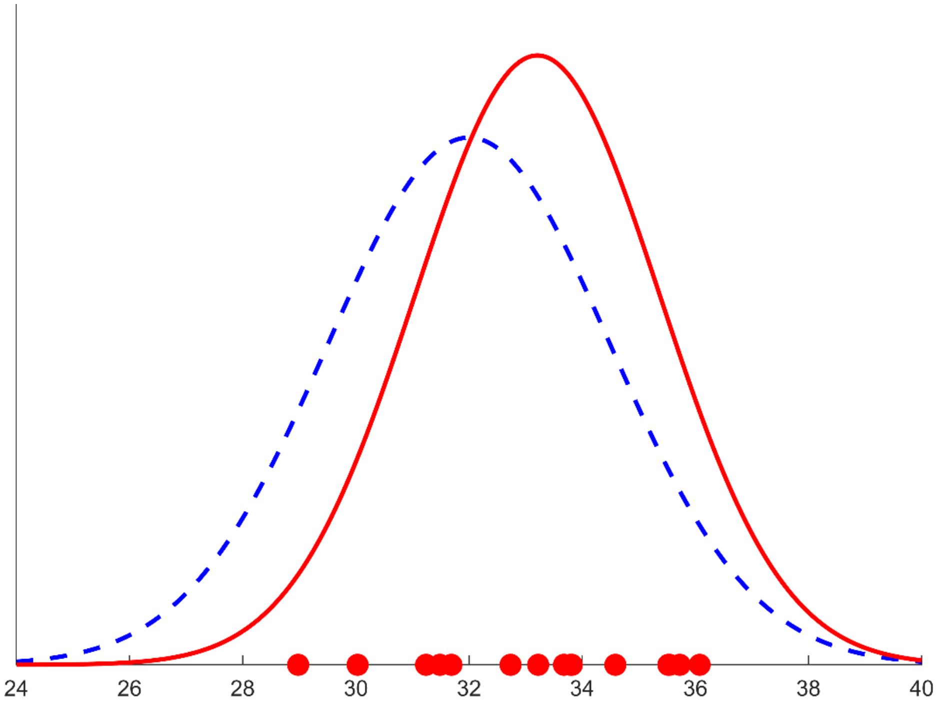


ako je $x_1 < x_2$ onda je $\ln x_1 < \ln x_2$

To znači da je vrednost θ koja maksimizuje ln *L*(θ) takođe i vrednost koja maksimizuje *L*(θ)

- Maksimizacija logaritma ekvivalentna je minimizaciji negativnog logaritma
- **Analitička prednost**: $\log x \cdot y = \log x + \log y$ → lakše ćemo pronaći izvod
- **Numerička prednost**: umesto računanja proizvoda mnogo malih brojeva (verovatnoća) što može dovesti do *underflow*-a numeričke preciznosti računara, računaćemo zbir logaritama tih brojeva

ML primer



$$L(\mu, \sigma^2) = p(x|\mu, \sigma^2) =$$

$$\prod_{n=1}^{N} \mathcal{N}(x^{(n)}|\mu, \sigma^2)$$

- Imamo uzorak $x = \left(x^{(1)}, \ldots, x^{(N)}\right)$
- ***Pretpostavka***: $x^{(i)}$ su uzorkovane nezavisno jedna od druge iz iste distribucije $\mathcal{N}\left(x^{(n)}\middle|\mu, \sigma^2\right)$
- ***Zadatak***: pronaći parametre modela μ i σ^2

ML primer

$$L(\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x^{(n)} | \mu, \sigma^2)$$

$\log x \cdot y = \log x + \log y$

$$\ln L(\mu, \sigma^2) = \sum_{n=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(x^{(n)} - \mu)^2}$$

$$\ln L(\mu, \sigma^2) = \sum_{n=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} + \sum_{n=1}^{N} \ln\left(\mathrm{e}^{-\frac{1}{2\sigma^2}(x^{(n)} - \mu)^2}\right)$$

$$\ln L(\mu, \sigma^2) = \sum_{n=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} + \sum_{n=1}^{N} -\frac{1}{2\sigma^2} (x^{(n)} - \mu)^2$$

ML primer

$$\ln L(\mu, \sigma^2) = \sum_{n=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} + \sum_{n=1}^{N} -\frac{1}{2\sigma^2} (x^{(n)} - \mu)^2$$

$\log x/y = \log x - \log y$

$$\ln L(\mu, \sigma^2) = \sum_{n=1}^{N} -\ln(2\pi\sigma^2)^{\frac{1}{2}} + \sum_{n=1}^{N} -\frac{1}{2\sigma^2} (x^{(n)} - \mu)^2$$

$\log x^y = y \log x$

$$\ln L(\mu, \sigma^2) = -\frac{N}{2} ln\sigma^2 - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x^{(n)} - \mu)^2$$

Određivanje μ

$$\ln L(\mu, \sigma^2) = -\frac{N}{2} ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x^{(n)} - \mu)^2$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{-2\sum_{n=1}^{N}(x^{(n)} - \mu)(-1)}{2\sigma^2} = 0$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x^{(n)}$$

Određivanje σ

$$\ln L(\mu, \sigma^2) = -\frac{N}{2} ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x^{(n)} - \mu)^2$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{-N}{2\sigma^2} + \frac{\sum_{n=1}^{N}\left(x^{(n)} - \mu\right)^2}{2\sigma^2} = 0$$

$$\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x^{(n)} - \mu_{ML})^2$$