MLE method

Maximum Likelihood Estimate (MLE)is **a method of estimating the parameters of an assumed probability distribution, given some observed data**. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable.

Maximum Likelihood Estimate (MLE) on menetelmä oletetun todennäköisyysjakauman parametrien arvioimiseksi käytettävissä olevan havaintodatan perusteella. Tämä saavutetaan maksimoimalla ns. uskottavuusfunktio L (Likelihood function) siten, että oletetun tilastollisen mallin mukaan havaittu data on todennäköisin.

Merkinnät:
- Olkoon x satunnaismuuttuja ja f(x,a,b) sen tiheysfunktio, missä a ja b ovat jakaumaparametrit (esim. a=keskiarvo ja b=keskihajonta (useissa jakaumissa jakaumaparametrit ovat muut kuin keskiarvo ja keskihajonta)).

- Olkoot $x_1, x_2, \ldots, x_n$ havaitut muuttujan x arvot.

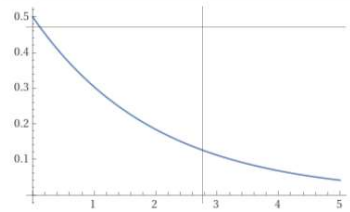
Todennäköisyys sille, että havainto x_i on peräisin oletetusta jakaumasta on $f(x_i,a,b)$. (Tämä on itse asiassa tiheysfunktion f määritelmä)

Tuloperiaatteen nojalla todennäköisyys sille, että koko havaintosarja $x_1, x_2, \ldots, x_n$ on peräisin jakaumasta f(x,a,b) = Likelihood funktio L (suomeksi uskottavuusfunktio)

$$L = \prod_{i=1}^{n} f(x_i, a, b)$$

L on siis funktio, jonka muuttujina ovat jakaumaparametrit a ja b. MLE menetelmässä etsitään uskottavuusfunktion L minimikohta. Tuloksena saadaan jakaumaparametreille a ja b arvot.

Esim1. Eksponenttijakauma on yksiparametrinen, yksinkertainen todennäköisyysjakauma, jota käytetään esimerkiksi hehkulamppujen vikaantumisaikojen mallinnuksessa.
Eksponenttijakauman tiheysfunktio $f(x,\lambda) = \lambda * e^{-\lambda x}$, missä $x \geq 0$.



Testissä oli 6 hehkulamppua, jotka rikkoontuivat ajoissa (1.1, 1.35, 1.5, 1.75, 1.8 ja 1.97) * 1000 h.
Määritä tiheysfunktio f(x,λ) eli parametrin λ arvo MLE menetelmällä.

Maksimoitava funktio $L = \prod \lambda * e^{-\lambda xi} = \lambda^6 e^{-\lambda(x1+x2+...+x)} = \lambda^6 e^{-\lambda * 9.47}$

Maksimiarvo löytyy parametrin λ arvolla 0.6336

Vastaus: $f(x) = 0.6336 * e^{-0.6336x}$
! Tehtävä voidaan ratkaista helpommin käyttäen tietoa, että eksponenttijakauman f(x, λ) keskiarvo on 1/ λ.
Otoskeskiarvo luvuista (1.1, 1.35, 1.5, 1.75, 1.8 ja 1.97) on 1.5783 ,
joten parametri λ = 1/1.5783 = 0.6336.

Esim2. Suomalaisen ampumahiihtäjän pystyammuntapaikalla pudotettujen taulujen lukumäärät kevään MC kiertueella olivat 3,4,4,5,4,5,3,3,4,4. Oletetaan, lukumäärät noudattavat binomijakaumaa, jossa yksittäiseen tauluun osumisen todennäköisyys p on vakio.
Määritä p:n arvo perustuen dataan. Käytä MLE menetelmää.

Maksimoitava uskottavuusfunktio olisi

$$L = \prod_{i=1}^{10} \binom{5}{x_i} p^{x_i} (1-p)^{5-x_i}$$

Funktio on muodoltaan sellainen, että sen ääriarvokohdan määritys on vaikeaa. Ottamalla funktiosta L logaritmi, päästään funktioon LL, joka saavuttaa maksimiarvon samassa kohdassa.

$$LL = \sum_{i=1}^{10} (\binom{5}{x_i} + x_i ln(p) + (5-x_i) ln(1-p))$$

Sen derivaatta LL'(p) = ∑(0 +xi/p – (5-xi)/(1-p)) jonka nollakohta saadaan yhtälöstä
∑xi/p = ∑(5-xi)/(1-p) , josta ristiinkertomalla (1-p)∑xi = p∑(5-xi) , josta edelleen
∑xi = p(∑xi + ∑(5-xi)) josta saadaan p = ∑xi/((∑xi + ∑(5-xi)) = 39/(39+11) = 0.78.

! Tehtävän ratkaiseminen MLE:llä on tarpeetonta, koska tulos on itsestään selvä muutoinkin.
Urheilija pudotti taulun 39 laukauksella 50 yrityksestä, joten P(osuu tauluun) = 39/50 = 0.78.

Esim3. Gaussin jakauman tiheysfunktio on

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

Määritä MLE menetelmällä jakaumaparametrit μ ja σ perustuen seuraaviin satunnaismuuttujan x havaintoarvoihin:
175, 178, 184, 183, 181, 176, 185.

Likelihood-funktio on tässä tapauksessa seuraava:

$$L = \prod_{i=1}^{7} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Tässäkin tapauksessa on helpompaa käyttää logaritmista ln(L) – funktiota

$$LL = \sum_{i=1}^{n} (ln \frac{1}{\sqrt{2\pi\sigma^2}} + - \frac{(x_i-\mu)^2}{2\sigma^2})$$
$$= \sum_{i=1}^{n} (ln \frac{1}{\sqrt{2\pi}} - ln\sigma - \frac{(x_i-\mu)^2}{2\sigma^2})$$

Minimi löytyy kohdasta, jossa osittaisderivaatat ovat nollia:
$dLL/d\mu = 2/2\sigma^2 \sum(xi-\mu)=0 \Rightarrow \sum xi = n\,\mu \Rightarrow \mu = \sum xi/n$ (= otoskeskiarvo, kuten voi olettaa)

$dLL/d\sigma = \sum(-1/\sigma + 2(xi-\mu)^2/2\,\sigma^3) = 0 \Rightarrow \sum(-\sigma^2 + (xi-\mu)^2) = 0 \Rightarrow -7\sigma^2 + \sum(xi-\mu)^2 = 0$
$\Rightarrow$ varianssi $\sigma^2 = \sum(xi-\mu)^2/n$, josta neliöjuuren otolla saadaan keskihajonta σ

Esimerkissä n = 7 ja ∑xi = 1262 => keskiarvo μ = 1262/7 = 180.3
Keskihajonnaksi tulee vastaavasti 4.0
Tässäkään esimerkissä MLE metodista ei ole varsinaista hyötyä, tulokset olisi saatu muutenkin helpommalla tavalla.

Logistinen regressio

= yksi koneoppimisen perusmenetelmiä, jota käytetään luokitteluun

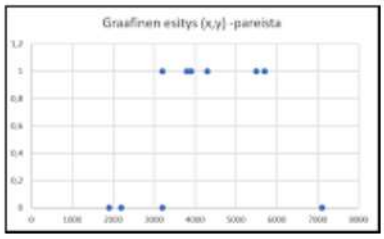
Esim4. **Riippumaton muuttuja välimatka-asteikollinen,**
Riippuva muuttuja on 0 – 1 muuttuja

Logistinen regressio muodostaa datan perusteella funktion p(x), jonka arvojoukko on [0, ∞ [.
Funktion p(x) arvo = todennäköisyys, että selitettävä muuttuja saa arvon 1 kyseisellä x:n arvolla.

Esimerkkidata

KK-tulot = X	Harkitsee sähköautoa = p
2200	0
3200	1
1900	0
5500	1
4300	1
3800	0
7100	1
5700	1
3200	0
3900	1

Datan graafinen esitys

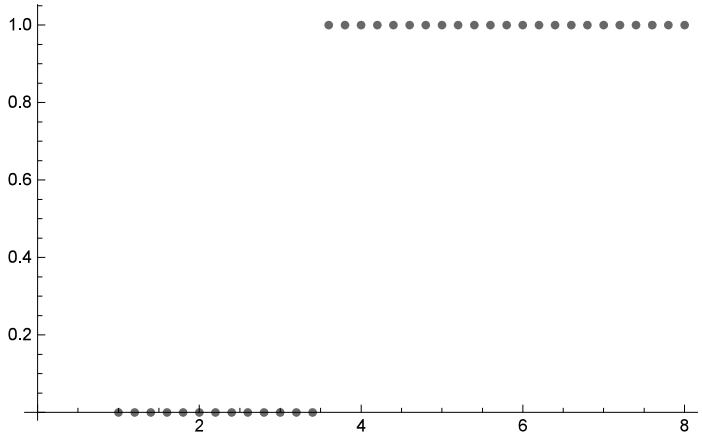


Datamuuttuja

Kk palkka muuttujan x arvot muunnettu yksikköön k€

```
data = {{2.2, 0}, {3.2, 1}, {1.9, 0}, {5.5, 1},
{4.3, 1}, {3.8, 0}, {7.1, 1}, {5.7, 1}, {3.2, 0}, {3.9, 1}};
```

```
ListPlot[ennusteet]
```



Logistisen regression teoriaa, manuaaliratkaisu

Sigmoidimallin laskeminen manuaalisesti

KK-tulot = X	Harkitsee sähköautoa = p
2200	0
3200	1
1900	0
5500	1
4300	1
3800	0
7100	1
5700	1
3200	0
3900	1

Muuttujan vaihdos

- Logistinen regressio on sukua lineaariselle regressiolle. Erona on se, että selitettävällä muuttujalla p on vain
kaksi mahdollista arvoa: 0 ja 1.
Lineaarisessa regressiossa selitettävän muuttujan arvoalue on]-∞, ∞ [

- Logistisessa regressiossa selitettävälle muuttujalle tehdään muunnos, jolla sen arvoalueeksi tulee]-∞, ∞ [

$$y = \ln(\text{odds}[p]) = \ln\left[\frac{p}{1-p}\right] \tag{1}$$

Vedonlyöntikertoimiin liittyvä funktio odds(p)

Funktiota odds(p)= $\frac{p}{1-p}$ sanotaan vedonlyöntikerroinfunktioksi tai riskifunktioksi. Esimerkiksi jos ravihevosen

voiton todennäköisyydeksi arvioidaan 75%, niin odds(p)= $\frac{0.75}{1-0.75}$ = 3, eli vedonlyöntikielellä 3:1.

Logistisessa regressiossa todennäköisyysmuuttuja muunnetaan vedonlyöntifunktion luonnolliseksi logaritmiksi.

Data esitettynä uusilla muuttujilla (x,y)

Data-taulukko muuttuu siten, että arvoa p = 1 vastaa arvo y =$\ln\left[\frac{1}{0}\right]=\infty$,
arvoa p=0 vastaa y =$\ln\left[\frac{0}{1}\right]=-\infty$ ja arvoa p = 0.5 vastaa arvo y =$\ln\left[\frac{0.5}{0.5}\right]=0$

KK-tulot = X	Harkitsee sähköautoa = p	y=logit/(p) =log(p/(1-p))
2200	0	-∞
3200	1	∞
1900	0	-∞
5500	1	∞
4300	1	∞
3800	0	∞
7100	1	-∞
5700	1	∞
3200	0	-∞
3900	1	∞

Lineaarisen mallin y = a x + b sovitus dataan.

Tavanomaisessa lineaarisessa regressiossa haetaan minimi residuaalivektorin pituuden neliölle, eli siinä minimoidaan
havaittujen y - arvojen ja mallin avulla laskettujen y - arvojen erotusten neliösumma $\sum(y_i - ax_i - b)^2$

Koska taulukon datassa y - arvot ovat äärettömän suuria, ei neliösummaa ole mahdollista muodostaa.
Tarvitaan toinen menetelmä käyrän sovitukseen data-arvoihin: Maximum Likelihood Method

Maximum Likelihood Method = Suurimman Todennäköisyyden Menetelmä

Käänteismuunnos y → p

Tehdään uudelle muuttujalle käänteismuunnos ja palataan alkuperäiseen todennäköisyysfunktioon p(x).
y=ln($\frac{p}{1-p}$) → p = e^y(1-p) → p(1+e^y) =e^y → p =$\frac{e^y}{1+e^y}$ = $\frac{1}{1+e^{-y}}$

$$p(x) = \frac{1}{1 + e^{-y}} \tag{2}$$

Sijoitetaan y = a x + b ja saadaan

$$p(x) = \frac{1}{1 + e^{-(a x+b)}} \tag{3}$$

Suurimman todennäköisyyden antavien parametrien laskeminen

Funktion p(x) tulkinnasta seuraa, että
(i) todennäköisyys sille, että selitettävä muuttuja saa arvon 1 x:n arvolla x_i on $p(x_i)$
(ii) todennäköisyys sille, että selitettävä muuttuja saa arvon 0 x:n arvolla x_i on 1 - $p(x_i)$

Lasketaan todennäköisyys sille, että havaitaan edellä annettu datajoukko. Tämä todennäköisyys on tulo

$$L = \left(1 - \frac{1}{1 + e^{-(a\,2.2+b)}}\right)\left(\frac{1}{1 + e^{-(a\,3.2+b)}}\right)$$
$$\left(1 - \frac{1}{1 + e^{-(a\,1.9+b)}}\right)\left(\frac{1}{1 + e^{-(a\,5.5+b)}}\right)\left(\frac{1}{1 + e^{-(a\,4.3+b)}}\right)\left(1 - \frac{1}{1 + e^{-(a\,3.8+b)}}\right)$$
$$\left(\frac{1}{1 + e^{-(a\,7.1+b)}}\right)\left(\frac{1}{1 + e^{-(a\,5.7+b)}}\right)\left(1 - \frac{1}{1 + e^{-(a\,3.2+b)}}\right)\left(\frac{1}{1 + e^{-(a\,3.9+b)}}\right);$$

Voisimme laskea funktion L maksimia vastaavat parametrien a ja b arvot gradientin nollakohdasta. Tulomuodosta johtuen maksimin löytäminen voisi olla laskennallisesti vaikeaa. Siksi otetaan luonnollinen logaritmi funktiosta L , jolloin saadaan yksinkertaisempi funktio LL (logarimic likelihood). Logaritmin laskusääntöjä käyttäen funktio LL näyttää seuraavalta:

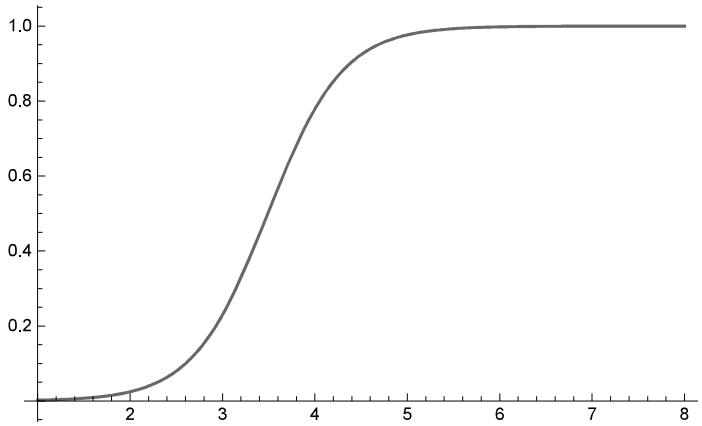
```
(*Logarithmic Likelihood Function:*)
```
$$LL = \text{Log}\left[1 - \frac{1}{1 + e^{-(a\,2.2+b)}}\right] - \text{Log}\left[1 + e^{-(a\,3.2+b)}\right] + \text{Log}\left[1 - \frac{1}{1 + e^{-(a\,1.9+b)}}\right] -$$
$$\text{Log}\left[1 + e^{-(a\,5.5+b)}\right] - \text{Log}\left[1 + e^{-(a\,4.3+b)}\right] + \text{Log}\left[1 - \frac{1}{1 + e^{-(a\,3.8+b)}}\right] -$$
$$\text{Log}\left[1 + e^{-(a\,7.1+b)}\right] - \text{Log}\left[1 + e^{-(a\,5.7+b)}\right] + \text{Log}\left[1 - \frac{1}{1 + e^{-(a\,3.2+b)}}\right] - \text{Log}\left[1 + e^{-(a\,3.9+b)}\right];$$

Haetaan maksimiarvo LL funktiolle

```
Maximize[LL, {a, b}]
```
```
{-3.16752, {a → 2.47077, b → -8.62413}}
```

Todennäköisyyttä kuvaava sigmoidi näyttää seuraavalta

```
Plot[1/(1 + e^-(2.47 x-8.62)), {x, 1, 8}]
```



Maksimi näkyy myös LL funktion Contour diagrammissa vaaleana alueena.

```
ContourPlot[LL, {a, 1, 7}, {b, -20, 0}]
```