J.T. 2024

Maximum Likelihood menetelmä MLE
(suom.: suurimman uskottavuuden menetelmä)

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, …., 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, …., 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 maksimikohta. Tuloksena saadaan jakaumaparametreille a ja b arvot.

Monissa tapauksissa uskottavuusfunktio L sisältää potenssilausekkeiden tuloja ja osamääriä, sekä mahdollisesti eksponenttifunktioita. Tällöin uskottavuusfunktion *L* sijasta haetaan maksimia sen logaritmille ln(L).

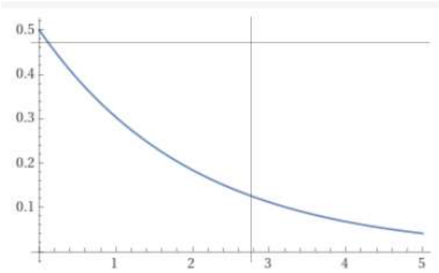
$$LL = \ln(L)$$

Perustelu: Jos g(x) on aidosti kasvava funktio, niin funktio g(f(x)) saavuttaa maksimiarvon kohdassa, jossa f(x):llä on maksimi. Funktion ln(x) on aidosti kasvava, joten ln(f(x)) saavuttaa maksimiarvon kohdassa, jossa f(x):llä on maksimi

----

J.T. 2024

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 $f(x,\lambda)$ eli parametrin λ arvo MLE menetelmällä.

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

Funktio on sellaista muoto, että on helpompi maksimoida sen logaritmi

$LL = 6 \ln(\lambda) - 9.47 \lambda$ (*käytetyt säännöt liite1:ssä)

Maksimi on derivaatan nollakohdassa: $6/\lambda - 9.47 = 0 \Rightarrow \lambda = 6/9.47 = 0.6336$

Vastaus: Kysytty tiheysfunktio $f(x) = 0.6336 * e^{-0.6336 x}$

! Edellinen tehtävä voidaan ratkaista helpommin käyttäen tietoa, että eksponenttijakauman $f(x, \lambda)$ keskiarvo on $1/\lambda$. Otoskeskiarvo luvuista (1.1, 1.35, 1.5, 1.75, 1.8 ja 1.97) on 1.5783 ,
joten parametri $\lambda = 1/1.5783 = 0.6336$.

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 on seuraava:

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

Funktio on muodoltaan sellainen, että sen ääriarvokohdan määritys on helpompi suorittaa käyttämällä logaritmia ln(L), 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 $\partial LL/\partial p = \sum(0 + x_i/p - (5-x_i)/(1-p))$ jonka nollakohta saadaan yhtälöstä
$\sum x_i/p = \sum(5-x_i)/(1-p)$, josta ristiinkertomalla $(1-p)\sum x_i = p\sum(5-x_i)$, josta edelleen
$\sum x_i = p(\sum x_i + \sum(5-x_i))$ josta saadaan $p = \sum x_i/((\sum x_i + \sum(5-x_i)) = 39/(39+11) = 0.78$.

! Tehtävän ratkaiseminen MLE:llä on turhan monimutkainen tapa, 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:
$\partial LL/\partial\mu = 2/2\sigma^2 \sum(x_i-\mu)=0 \Rightarrow \sum x_i = n\,\mu \Rightarrow \mu = \sum x_i/n$ (= otoskeskiarvo, kuten voi olettaa)

$\partial LL/\partial\sigma = \sum(-1/\sigma + 2(x_i-\mu)^2/2\,\sigma^3) = 0 \Rightarrow \sum(-\sigma^2 + (x_i-\mu)^2) = 0 \Rightarrow -7\sigma^2 + \sum(x_i-\mu)^2 = 0$
=> varianssi $\sigma^2 = \sum(x_i-\mu)^2/n$, jonka neliöjuuri on keskihajonta σ

Esimerkissä n = 7 ja $\sum x_i = 1262$ => 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:
”Parhaat estimaatit populaatiokeskiarvolle, ja – keskihajonnalle ovat otoskeskiarvo ja otoskeskihajonta”

J.T. 2024

Logistinen regressio

- on yksi koneoppimisen perusmenetelmiä.

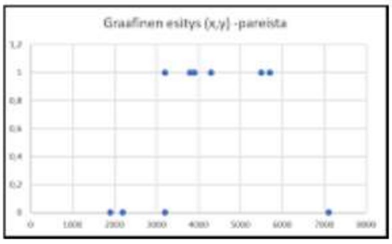
Siinä riippumaton muuttuja on välimatka-asteikollinen ja riippuva muuttuja on 0,1 -muuttuja. . Logistisen regression menetelmässä muodostetaan datan perusteella funktion p(x), jonka arvojoukko on [0, ∞[. Funktion p(x) arvo = todennäköisyys, että selitettävä muuttuja saa arvon 1 kyseisellä x:n arvolla.

Esim4. Selittävä muuttuja x = henkilön kk-tulot, selitettävä muuttuja y on 0,1 -muuttuja ”harkitsee sähköauton ostoa”. Alla on dataa 10 henkilön mitatuista (x,y) -arvoista.

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



Havaintoaineisto pistepareina:

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]];
```

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]-∞, ∞ [

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

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	∞
3900	0	-∞
5500	1	∞
4300	1	∞
3800	0	∞
7100	1	-∞
5700	1	∞
3300	0	∞
3800	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 $\sum (y_i - a x_i - b)^2$

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).
$y = \ln\left(\frac{p}{1-p}\right) \rightarrow\ p = e^y(1-p) \rightarrow p(1+e^y) = e^y \rightarrow p = \frac{e^y}{1+e^y} = \frac{1}{1+e^{-y}}$

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

Sijoitetaan y = a x + b ja saadaan

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

Suurimman uskottavuuden antavien parametrien laskeminen.

Funktion p(x) tulkinnasta seuraa, että

(i) todennäköisyys sille, että selitettävä muuttuja saa arvon 1 x:n arvoilla x_i on $p(x_i)$

(ii) todennäköisyys sille, että selitettävä muuttuja saa arvon 0 x:n arvoilla x_i on $1 - p(x_i)$

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

$$L = \left(1-\frac{1}{1+e^{-(a\,2.3+b)}}\right)\left(\frac{1}{1+e^{-(a\,3.2+b)}}\right)\left(1-\frac{1}{1+e^{-(a\,1.9+b)}}\right)\left(\frac{1}{1+e^{-(a\,5.5+b)}}\right)\left(\frac{1}{1+e^{-(a\,4.3+b)}}\right)\left(1-\frac{1}{1+e^{-(a\,2.8+b)}}\right)\left(\frac{1}{1+e^{-(a\,7.1+b)}}\right)\left(\frac{1}{1+e^{-(a\,5.7+b)}}\right)\left(1-\frac{1}{1+e^{-(a\,3.2+b)}}\right)\left(\frac{1}{1+e^{-(a\,3.9+b)}}\right);$$

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:*)

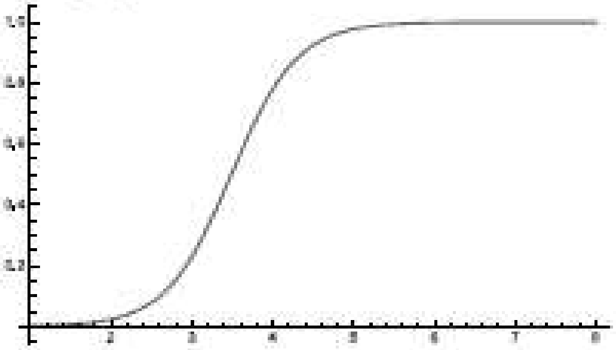
$$LL = \mathrm{Log}\left[1-\frac{1}{1+e^{-(a\,2.3+b)}}\right]-\mathrm{Log}\left[1+e^{-(a\,3.2+b)}\right]+\mathrm{Log}\left[1-\frac{1}{1+e^{-(a\,1.9+b)}}\right]-\mathrm{Log}\left[1+e^{-(a\,5.5+b)}\right]-\mathrm{Log}\left[1+e^{-(a\,4.3+b)}\right]+\mathrm{Log}\left[1-\frac{1}{1+e^{-(a\,2.8+b)}}\right]-\mathrm{Log}\left[1+e^{-(a\,7.1+b)}\right]-\mathrm{Log}\left[1+e^{-(a\,5.7+b)}\right]+\mathrm{Log}\left[1-\frac{1}{1+e^{-(a\,3.2+b)}}\right]-\mathrm{Log}\left[1+e^{-(a\,3.9+b)}\right];$$

Haetaan maksimiarvo LL funktiolle

```
Maximize[LL, {a, b}]
{-3.16752, {a -> 2.47877, b -> -8.62413}}
```

Todennäköisyyttä kuvaava sigmoidi näyttää seuraavalta

$$\mathrm{Plot}\left[\frac{1}{1+e^{-(2.47\,x-8.62)}},\ \{x, 1, 8\}\right]$$



Maksimi näkyy myös LL funktion Contour diagrammissa vaaleana alueena.

Tulkinta: Jos tulot ylittävät 3500 eur/kk, todennäköisyys sähköauton ostoon on yli 50%.

Alla on sama tehtävä suoritettu Mathematica- ohjelmiston valmisfunktiolla LogitModelFit. Tulos on sama kuin yllä kuvatussa manuaaliratkaisussa.

```
malli = LogitModelFit[data, x, x]
```

FittedModel$\left[\boxed{\frac{1}{1+e^{8.62413-2.47077x}}} \right]$

Malli funktiomuodossa

```
Normal[malli]
```

$$\frac{1}{1+e^{8.62413-2.47077\,x}}$$

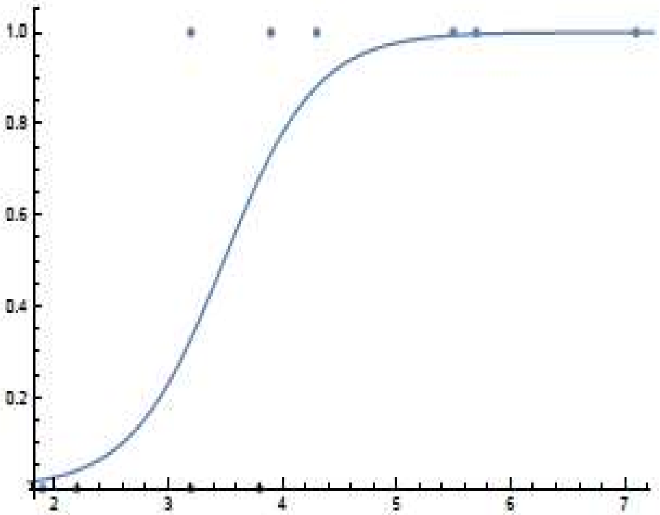
Lasketaan mallia käyttäen todennäköisyys kohdassa x =3.7

```
malli[3.7]
```

0.626618

Kuva alkuperäisestä datasta ja siihen sovitetusta sigmoidista.

```
Show[ListPlot[data], Plot[malli[x], {x, 1.0, 7.5}]]
```



Lasketaan x:n arvo, jossa todennäköisyys ylittää 0.5.

```
Solve[malli[x] == 0.5, x]
```

$\{\{x \to 3.49046\}\}$

Tulkinta: Jos kk-tulot ylittävät 3500 Euroa, on todennäköistä, että henkilö harkitsee sähköauton

LIITE1: Esimerkeissä käytettyjä laskusääntöjä:

$e^a e^b = e^{a+b}$

$\ln(x*y) = \ln(x) + \ln(y)$
$\ln(x/y) = \ln(x) – \ln(y)$
$\ln(x^r) = r \ln(x)$

Derivointikaavoja:

$Dx^n = n x^{n-1}$
$D(1/x^n) = -n/x^{n+1}$
$De^x = e^x$
$De^{ax} = a e^{ax}$
$D\ln(x) = 1/x$