

Regressziós modellek alternatív becslési lehetőségei

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 9.

Tartalom

1 A maximum likelihood (ML) elv

Tartalom

1 A maximum likelihood (ML) elv

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A likelihood fogalma

- Likelihood: folytonos változónál a sűrűségfüggvény helyettesítési értéke adott ponton, diszkrétnél a valószínűség (tehát a valószínűségi súlyfüggvény értéke adott ponton)
- (Lehet többdimenziós is)
- Nem mondhatunk valószínűséget helyette, hiszen folytonos esetben nem valószínűségről van szó (csak gondoljunk bele, a sűrűség simán lehet 1-nél nagyobb)
- Részletesebben: folytonos változó minden adott konkrét értéket nulla valószínűséggel vesz fel!
- De a likelihood mégis bír kézzelfogható értelmezéssel: ezzel arányos valószínűséggel esik a valószínűségi változó az adott pont kis környezetébe
- (A '80-as években kísérleteztek a „valószerűség” szó meghonosításával erre...)

A maximum likelihood (ML) elv

- Ha ismerjük a sokasági paraméterek értékét és egy modellt a mintavételre, akkor meg tudjuk mondani, hogy adott minta kijövetelének mekkora a likelihood-ja
- (Hiszen a modell épp ezt írja le, csak épp függ a sokasági paraméterektől, de ha azokat is ismerjük, akkor már egy konkrét számot kapunk)
- Az ML-becslés alapgondolata: ha nem ismerjük a sokasági paramétereket, akkor válasszuk azokat becslésnek, amely mellett a lehető legnagyobb a likelihood-ja annak, hogy az a minta jöjjön ki, ami ténylegesen ki is jött
- Józan paraszti ésszel is ésszerű, de nagyon fontos felhívni a figyelmet, hogy ez *nem* ugyanaz, mint hogy azt határozzuk meg, hogy mely paramétereknek a legnagyobb a likelihoodja (fordított a feltételezés iránya!)
- (Ez az igazán kézenfekvő kérdés, csak hogy ez igényel ismereteket arról, hogy a paraméterek eloszlása milyen még mielőtt egyáltalán mintát vettünk volna – itt válik el a frekvencionista és a bayes-i statisztika)

A maximum likelihood (ML) elv

- Ha ismerjük a sokasági paraméterek értékét és egy modellt a mintavételre, akkor meg tudjuk mondani, hogy adott minta kijövetelének mekkora a likelihood-ja
- (Hiszen a modell épp ezt írja le, csak épp függ a sokasági paraméterektől, de ha azokat is ismerjük, akkor már egy konkrét számot kapunk)
- Az ML-becsolás alapgondolata: ha nem ismerjük a sokasági paramétereket, akkor válasszuk azokat becsolásnak, amely mellett a lehető legnagyobb a likelihood-ja annak, hogy az a minta jöjjön ki, ami ténylegesen ki is jött
- Józan paraszti észszel is észszerű, de nagyon fontos felhívni a figyelmet, hogy ez *nem* ugyanaz, mint hogy azt határozzuk meg, hogy mely paramétereknek a legnagyobb a likelihoodja (fordított a feltételezés iránya!)
- (Ez az igazán kézenfekvő kérdés, csak hogy ez igényel ismeretet arról, hogy a paraméterek eloszlása milyen még mielőtt egyáltalán mintát vettünk volna – itt válik el a frekvencionista és a bayes-i statisztika)

A maximum likelihood (ML) elv

- Ha ismerjük a sokasági paraméterek értékét és egy modellt a mintavételre, akkor meg tudjuk mondani, hogy adott minta kijövetelének mekkora a likelihood-ja
- (Hiszen a modell épp ezt írja le, csak épp függ a sokasági paraméterektől, de ha azokat is ismerjük, akkor már egy konkrét számot kapunk)
- Az ML-becslés alapgondolata: ha nem ismerjük a sokasági paramétereket, akkor válasszuk azokat becslésnek, amely mellett a lehető legnagyobb a likelihood-ja annak, hogy az a minta jöjjön ki, ami ténylegesen ki is jött
- Józan paraszti ésszel is ésszerű, de nagyon fontos felhívni a figyelmet, hogy ez *nem* ugyanaz, mint hogy azt határozzuk meg, hogy mely paramétereknek a legnagyobb a likelihoodja (fordított a feltételezés iránya!)
- (Ez az igazán kézenfekvő kérdés, csak hogy ez igényel ismeretet arról, hogy a paraméterek eloszlása milyen még mielőtt egyáltalán mintát vettünk volna – itt válik el a frekvencionista és a bayes-i statisztika)

A maximum likelihood (ML) elv

- Ha ismerjük a sokasági paraméterek értékét és egy modellt a mintavételre, akkor meg tudjuk mondani, hogy adott minta kijövetelének mekkora a likelihood-ja
- (Hiszen a modell épp ezt írja le, csak épp függ a sokasági paraméterektől, de ha azokat is ismerjük, akkor már egy konkrét számot kapunk)
- Az ML-becslés alapgondolata: ha nem ismerjük a sokasági paramétereket, akkor válasszuk azokat becslésnek, amely mellett a lehető legnagyobb a likelihood-ja annak, hogy az a minta jöjjön ki, ami ténylegesen ki is jött
- Józan paraszti ésszel is ésszerű, de nagyon fontos felhívni a figyelmet, hogy ez *nem* ugyanaz, mint hogy azt határozzuk meg, hogy mely paramétereknek a legnagyobb a likelihoodja (fordított a feltételezés iránya!)
- (Ez az igazán kézenfekvő kérdés, csak hogy ez igényel ismeretet arról, hogy a paraméterek eloszlása milyen még mielőtt egyáltalán mintát vettünk volna – itt válik el a frekvencionista és a bayes-i statisztika)

A maximum likelihood (ML) elv

- Ha ismerjük a sokasági paraméterek értékét és egy modellt a mintavételre, akkor meg tudjuk mondani, hogy adott minta kijövetelének mekkora a likelihood-ja
- (Hiszen a modell épp ezt írja le, csak épp függ a sokasági paraméterektől, de ha azokat is ismerjük, akkor már egy konkrét számot kapunk)
- Az ML-becsolás alapgondolata: ha nem ismerjük a sokasági paramétereket, akkor válasszuk azokat becsolásnak, amely mellett a lehető legnagyobb a likelihood-ja annak, hogy az a minta jöjjön ki, ami ténylegesen ki is jött
- Józan paraszti észszel is észszerű, de nagyon fontos felhívni a figyelmet, hogy ez *nem* ugyanaz, mint hogy azt határozzuk meg, hogy mely paramétereknek a legnagyobb a likelihoodja (fordított a feltételezés iránya!)
- (Ez az igazán kézenfekvő kérdés, csak hogy ez igényel ismeretet arról, hogy a paraméterek eloszlása milyen még mielőtt egyáltalán mintát vettünk volna – itt válik el a frekvencionista és a bayes-i statisztika)

Egyszerű várható érték becslés ML-elven

- Csináljuk meg újra ugyanazt az egyszerű példát amit az OLS-elvnél láttunk, de immár ML-elvű becsléssel!
- Emlékeztetőül a modellünk: $Y \sim \mathcal{N}(\mu, \sigma_0^2)$ (ahol σ_0^2 ismert) és erre van egy n elemű fae mintánk
- A valódi várható érték μ , a feltételezett várható értéket, ami végigfut majd az összes lehetséges értéken, jelölje m (és a legjobb érték, amit becslésként elfogadunk $\hat{\mu}$)
- Minta igazából most az n elem együtt, tehát a minta likelihood-ja ennek az n -dimenziós – folytonos sűrűségfüggvénynek a helyettesítési értéke

Egyszerű várható érték becslés ML-elven

- Csináljuk meg újra ugyanazt az egyszerű példát amit az OLS-elvnél láttunk, de immár ML-elvű becsléssel!
- Emlékeztetőül a modellünk: $Y \sim \mathcal{N}(\mu, \sigma_0^2)$ (ahol σ_0^2 ismert) és erre van egy n elemű fae mintánk
- A valódi várható érték μ , a feltételezett várható értéket, ami végigfut majd az összes lehetséges értéken, jelölje m (és a legjobb érték, amit becslésként elfogadunk $\hat{\mu}$)
- Minta igazából most az n elem együtt, tehát a minta likelihood-ja ennek az n -dimenziós – folytonos sűrűségfüggvénynek a helyettesítési értéke

Egyszerű várható érték becslés ML-elven

- Csináljuk meg újra ugyanazt az egyszerű példát amit az OLS-elvnél láttunk, de immár ML-elvű becsléssel!
- Emlékeztetőül a modellünk: $Y \sim \mathcal{N}(\mu, \sigma_0^2)$ (ahol σ_0^2 ismert) és erre van egy n elemű fae mintánk
- A valódi várható érték μ , a feltételezett várható értéket, ami végigfut majd az összes lehetséges értékén, jelölje m (és a legjobb érték, amit becslésként elfogadunk $\hat{\mu}$)
- Minta igazából most az n elem együtt, tehát a minta likelihood-ja ennek az n -dimenziós – folytonos sűrűségfüggvénynek a helyettesítési értéke

Egyszerű várható érték becslés ML-elven

- Csináljuk meg újra ugyanazt az egyszerű példát amit az OLS-elvnél láttunk, de immár ML-elvű becsléssel!
- Emlékeztetőül a modellünk: $Y \sim \mathcal{N}(\mu, \sigma_0^2)$ (ahol σ_0^2 ismert) és erre van egy n elemű fae mintánk
- A valódi várható érték μ , a feltételezett várható értéket, ami végigfut majd az összes lehetséges értékén, jelölje m (és a legjobb érték, amit becslésként elfogadunk $\hat{\mu}$)
- Minta igazából most az n elem együtt, tehát a minta likelihood-ja ennek az n -dimenziós – folytonos sűrűségfüggvénynek a helyettesítési értéke

Egyszerű várható érték becslés ML-elven

- Szerencsére fae esetben ez szétesik szorzattá:

$$\begin{aligned}
 L_m(\mathbf{y}) &= f_{\mathbf{Y}}(\mathbf{y}) = f_m(\mathbf{y}) = \prod_{i=1}^n f_m(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot e^{-\frac{(y_i-m)^2}{2\sigma_0^2}} = \\
 &= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \cdot e^{-\sum_{i=1}^n \frac{(y_i-m)^2}{2\sigma_0^2}}
 \end{aligned}$$

- Célszerűbb ennek a maximalizálása helyett inkább a logaritmusát maximalizálni (a logaritmus monoton transzformáció, így a logaritmált maximum ugyanott van, mint az eredeti maximuma):

$$l_m(\mathbf{y}) = \log L_m(\mathbf{y}) = -\frac{n}{2} \log(2\pi\sigma_0^2) + \sum_{i=1}^n -\frac{(y_i-m)^2}{2\sigma_0^2}$$

Egyszerű várható érték becslés ML-elven

- Szerencsére fae esetben ez szétesik szorzattá:

$$\begin{aligned} L_m(\mathbf{y}) = f_{\mathbf{Y}}(\mathbf{y}) = f_m(\mathbf{y}) &= \prod_{i=1}^n f_m(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} \cdot e^{-\frac{(y_i-m)^2}{2\sigma_0^2}} = \\ &= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \cdot e^{-\sum_{i=1}^n \frac{(y_i-m)^2}{2\sigma_0^2}} \end{aligned}$$

- Célszerűbb ennek a maximalizálása helyett inkább a logaritmusát maximalizálni (a logaritmus monoton transzformáció, így a logaritmált maximum ugyanott van, mint az eredeti maximuma):

$$l_m(\mathbf{y}) = \log L_m(\mathbf{y}) = -\frac{n}{2} \log(2\pi\sigma_0^2) + \sum_{i=1}^n -\frac{(y_i-m)^2}{2\sigma_0^2}$$

Egyszerű várható érték becslés ML-elven

- A maximalizáláshoz deriváljuk m szerint:

$$\frac{d l_m(\mathbf{y})}{d m} = -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1)$$

- Tegyük egyenlővé nullával és oldjuk meg:

$$\begin{aligned} -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1) &= 0 \Rightarrow \sum_{i=1}^n 2(y_i - m) = 0 \\ \Rightarrow \widehat{\mu_{\text{ML}}} &= \frac{\sum_{i=1}^n y_i}{n} \end{aligned}$$

- (A második derivált pedig $-\frac{n}{2\sigma_0^2}$, negatív, tehát ez tényleg szélsőérték, és tényleg maximum)

Egyszerű várható érték becslés ML-elven

- A maximalizáláshoz deriváljuk m szerint:

$$\frac{d l_m(\mathbf{y})}{d m} = -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1)$$

- Tegyük egyenlővé nullával és oldjuk meg:

$$\begin{aligned} -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1) &= 0 \Rightarrow \sum_{i=1}^n 2(y_i - m) = 0 \\ \Rightarrow \widehat{\mu_{\text{ML}}} &= \frac{\sum_{i=1}^n y_i}{n} \end{aligned}$$

- (A második derivált pedig $-\frac{n}{2\sigma_0^2}$, negatív, tehát ez tényleg szélsőérték, és tényleg maximum)

Egyszerű várható érték becslés ML-elven

- A maximalizáláshoz deriváljuk m szerint:

$$\frac{d l_m(\mathbf{y})}{d m} = -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1)$$

- Tegyük egyenlővé nullával és oldjuk meg:

$$\begin{aligned} -\frac{1}{2\sigma_0^2} \sum_{i=1}^n 2(y_i - m) \cdot (-1) &= 0 \Rightarrow \sum_{i=1}^n 2(y_i - m) = 0 \\ \Rightarrow \widehat{\mu_{\text{ML}}} &= \frac{\sum_{i=1}^n y_i}{n} \end{aligned}$$

- (A második derivált pedig $-\frac{n}{2\sigma_0^2}$, negatív, tehát ez tényleg szélsőérték, és tényleg maximum)

Az ML-becslés általában

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} L_{\mathbf{t}}(\mathbf{y})$$

- Ha a mintánk fae, akkor:

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} \sum_{i=1}^n l_{\mathbf{t}}(y_i)$$

- Az ML-becslés valószínűségi modellt igényel (az OLS-becslésnél ez nem volt így, igazából elég volt az, hogy legyen egy predikált értékünk, az kijöhetett bárhogy), de cserében egy sor kellemes tulajdonsággal bír
- (*Általában*, tehát pusztán amiatt, hogy ML-becslésről van szó: függetlenül attól, hogy mi a probléma, ezeket a tulajdonságokat pusztán az ML-becslés mivolt miatt megkapjuk!)

Az ML-becslés általában

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} L_{\mathbf{t}}(\mathbf{y})$$

- Ha a mintánk fae, akkor:

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} \sum_{i=1}^n l_{\mathbf{t}}(y_i)$$

- Az ML-becslés valószínűségi modellt igényel (az OLS-becslésnél ez nem volt így, igazából elég volt az, hogy legyen egy predikált értékünk, az kijöhetett bárhogy), de cserében egy sor kellemes tulajdonsággal bír
- (*Általában*, tehát pusztán amiatt, hogy ML-becslésről van szó: függetlenül attól, hogy mi a probléma, ezeket a tulajdonságokat pusztán az ML-becslés mivolt miatt megkapjuk!)

Az ML-becslés általában

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} L_{\mathbf{t}}(\mathbf{y})$$

- Ha a mintánk fae, akkor:

$$\widehat{\theta}_{\text{ML}} = \arg \max_{\mathbf{t}} \sum_{i=1}^n l_{\mathbf{t}}(y_i)$$

- Az ML-becslés valószínűségi modellt igényel (az OLS-becslésnél ez nem volt így, igazából elég volt az, hogy legyen egy predikált értékünk, az kijöhetett bárhogy), de cserében egy sor kellemes tulajdonsággal bír
- (*Általában*, tehát pusztán amiatt, hogy ML-becslésről van szó: függetlenül attól, hogy mi a probléma, ezeket a tulajdonságokat pusztán az ML-becslés mivolt miatt megkapjuk!)

Az ML-beclés tulajdonságai

Jelesül egy ML-beclés elég általános körülmények között

- konzisztens
- aszimptotikusan torzítatlan
- aszimptotikusan hatásos
- aszimptotikusan normális
- invariáns ($g(\theta)$ ML-beclése $g(\widehat{\theta}_{\text{ML}})$, ha g egy kölcsönösen egyértelmű függvény)

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráeresztetni (hogyan megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráeresztelni (hogyan megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráereszteni (hogyan megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráeresztteni (hogyan megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráeresztelni (hogyan megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- De ezt az ML-elvet nem lehetne a sima lineáris modellre is ráereszteni (hogy megbecsüljük, de másképp mint eddig, OLS-elven)? Dehogynem!
- Egyetlen dologra van szükségünk: itt *mindenképp* fel kell tennünk valamit a hibatag eloszlásáról (különben az eredményváltozóra nem lesz eloszlásunk, anélkül pedig likelihood-unk sincs)
- Tegyük fel a normalitást (és a szferikális hibákat, tehát, hogy az $\underline{\varepsilon}$ kovarianciamátrixa $\sigma^2 \mathbf{I}$)
- Ekkor ugyanis \underline{Y} eloszlása is normális, mégpedig $\underline{X}\mathbf{b}$ várhatóértékkel és $\sigma^2 \mathbf{I}$ kovarianciamátrixszal (hiszen $\underline{Y} = \underline{X}\mathbf{b} + \underline{\varepsilon}$)
- Mivel azt mondtuk, hogy \mathbf{b} az (ismeretlen) sokasági paraméterek éppen feltételezett értéke; ebben fogunk majd maximalizálni
- Írhattuk volna azt is, hogy $Y_i = \underline{X}_i^T \mathbf{b} + \varepsilon_i$ és hozzátesszük, hogy a különböző megfigyelési egységek függetlenek

Lineáris modell becslése ML-elven

- A log-likelihood (a második megközelítésből felírva):

$$-\frac{n}{2} \log(2\pi\sigma^2) + \sum_{i=1}^n -\frac{(y_i - \underline{X}_i^T \mathbf{b})^2}{2\sigma^2}$$

- A log-likelihood (az első megközelítésből felírva, a többváltozós normális sűrűsége $\frac{1}{(2\pi)^{n/2} \det \mathbf{C}^{1/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$):

$$-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$$

- Ha ezt \mathbf{b} -ben maximalizáljuk, az ugyanaz, mint $(\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$ -t \mathbf{b} -ben minimalizálni
- ...de hát ez épp az amit az OLS-nél már megoldottunk!
- Úgyhogy az eredményt már tudjuk: $\widehat{\beta}_{\text{ML}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$

Lineáris modell becslése ML-elven

- A log-likelihood (a második megközelítésből felírva):

$$-\frac{n}{2} \log(2\pi\sigma^2) + \sum_{i=1}^n -\frac{(y_i - \underline{X}_i^T \mathbf{b})^2}{2\sigma^2}$$

- A log-likelihood (az első megközelítésből felírva, a többváltozós normális sűrűsége $\frac{1}{(2\pi)^{n/2} \det \mathbf{C}^{1/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$):

$$-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$$

- Ha ezt \mathbf{b} -ben maximalizáljuk, az ugyanaz, mint $(\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$ -t \mathbf{b} -ben minimalizálni
- ...de hát ez épp az amit az OLS-nél már megoldottunk!
- Úgyhogy az eredményt már tudjuk: $\widehat{\boldsymbol{\beta}}_{\text{ML}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$

Lineáris modell becslése ML-elven

- A log-likelihood (a második megközelítésből felírva):

$$-\frac{n}{2} \log(2\pi\sigma^2) + \sum_{i=1}^n -\frac{(y_i - \underline{X}_i^T \mathbf{b})^2}{2\sigma^2}$$

- A log-likelihood (az első megközelítésből felírva, a többváltozós normális sűrűsége $\frac{1}{(2\pi)^{n/2} \det \mathbf{C}^{1/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$):

$$-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \underline{\underline{X}}\mathbf{b})^T (\underline{y} - \underline{\underline{X}}\mathbf{b})$$

- Ha ezt \mathbf{b} -ben maximalizáljuk, az ugyanaz, mint $(\underline{y} - \underline{\underline{X}}\mathbf{b})^T (\underline{y} - \underline{\underline{X}}\mathbf{b})$ -t \mathbf{b} -ben minimalizálni
- ...de hát ez épp az amit az OLS-nél már megoldottunk!
- Úgyhogy az eredményt már tudjuk: $\widehat{\boldsymbol{\beta}}_{\text{ML}} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{y}$

Lineáris modell becslése ML-elven

- A log-likelihood (a második megközelítésből felírva):

$$-\frac{n}{2} \log(2\pi\sigma^2) + \sum_{i=1}^n -\frac{(y_i - \underline{X}_i^T \mathbf{b})^2}{2\sigma^2}$$

- A log-likelihood (az első megközelítésből felírva, a többváltozós normális sűrűsége $\frac{1}{(2\pi)^{n/2} \det \mathbf{C}^{1/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$):

$$-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \underline{\underline{X}}\mathbf{b})^T (\underline{y} - \underline{\underline{X}}\mathbf{b})$$

- Ha ezt \mathbf{b} -ben maximalizáljuk, az ugyanaz, mint $(\underline{y} - \underline{\underline{X}}\mathbf{b})^T (\underline{y} - \underline{\underline{X}}\mathbf{b})$ -t \mathbf{b} -ben minimalizálni
- ...de hát ez épp az amit az OLS-nél már megoldottunk!
- Úgyhogy az eredményt már tudjuk: $\widehat{\boldsymbol{\beta}}_{\text{ML}} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{y}$

Lineáris modell becslése ML-elven

- A log-likelihood (a második megközelítésből felírva):

$$-\frac{n}{2} \log(2\pi\sigma^2) + \sum_{i=1}^n -\frac{(y_i - \underline{X}_i^T \mathbf{b})^2}{2\sigma^2}$$

- A log-likelihood (az első megközelítésből felírva, a többváltozós normális sűrűsége $\frac{1}{(2\pi)^{n/2} \det \mathbf{C}^{1/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$):

$$-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$$

- Ha ezt \mathbf{b} -ben maximalizáljuk, az ugyanaz, mint $(\underline{y} - \underline{X}\mathbf{b})^T (\underline{y} - \underline{X}\mathbf{b})$ -t \mathbf{b} -ben minimalizálni
- ...de hát ez épp az amit az OLS-nél már megoldottunk!
- Úgyhogy az eredményt már tudjuk: $\widehat{\boldsymbol{\beta}}_{\text{ML}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$

Lineáris modell becslése ML-elven

- A lineáris regressziós modell esetén tehát az ML-becslés és az OLS-becslés egybeesik (ebből adódóan a tulajdonságaik is azonosak)
- De az ML-hez fel kellett tenni a hibanormalitást, az OLS-nél erre nem volt szükség

Lineáris modell becslése ML-elven

- A lineáris regressziós modell esetén tehát az ML-becslés és az OLS-becslés egybeesik (ebből adódóan a tulajdonságaik is azonosak)
- De az ML-hez fel kellett tenni a hibanormalitást, az OLS-nél erre nem volt szükség