

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Wahrscheinlichkeitsrechnung

Dr. rer. nat Dennis Müller

November 6, 2015

Table of Contents

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

1 Motivation und Überblick

2 Maximum Likelihood Methode

3 Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Parameterschätzung - Motivation und Überblick

- Viele Größen, die in praktischen Problemen auftauchen können entweder gar nicht oder nicht mit vertretbarem Aufwand gemessen werden.
- Beispiel: Wie schwer ist der Mond? Wie groß sind Menschen im Mittel? Wie weit ist ein anderes Fahrzeug von mir entfernt? Wieviel Wasser ist im Atlantik?
- Man kann jedoch sehr oft stichprobenartige Beobachtungen machen und dann versuchen, daraus auf die eigentlich interessante Größe zu schlußfolgern.
- Wir wollen diesen Prozeß im folgenden genauer betrachten und formalisieren.

Einführung

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Beispiel: Wir möchten wissen, wie groß Menschen im Mittel sind.
- Wir könnten alle Menschen (ca. 7 Mrd.) messen und den Mittelwert berechnen.
- Wir könnten auch nur 1000 Menschen messen und diesen Mittelwert bestimmen.
- Angenommen, dabei messen wir eine durchschnittliche Größe von 1.75m. Was wäre intuitiv die Schätzung für den Mittelwert der Gesamtheit?

- Wir betrachten im folgenden zwei verschiedene Techniken zur Parameterschätzung:
- Die so genannte

Maximum Likelihood Methode

- und die so genannte

Methode der kleinsten Quadrate.

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

**Maximum
Likelihood
Methode**

Methode der
kleinsten
Quadrate

Parameterschätzung - Maximum Likelihood Methode

Einführendes Beispiel

- Wir haben eine Urne mit roten und schwarzen Kugeln. Wir kennen das Verhältniss nicht und möchten dieses daher mit Hilfe einer Stichprobe schätzen.
- Wir ziehen 10 Kugeln mit zurücklegen und finden dabei 3 rote. Wir wissen bereits, das die Anzahl roter Kugeln einer Binomialverteilung mit unbekanntem Trefferwahrscheinlichkeit p folgt. Demnach ist die Wahrscheinlichkeit, dieses Ereigniss (3 rote unter 10 Kugeln) zu erhalten gegeben durch

$$P(A) = \binom{10}{3} p^3 (1-p)^7$$

- Intuitiv würden wir die Wahrscheinlichkeit für dieses Ereigniss gerne möglichst groß haben (weil es ja tatsächlich eingetreten ist). Wir suchen also das Maximum von $P(A)$ in Abhängigkeit von p .

Einführendes Beispiel

- Wir leiten $P(A)$ nach p ab und erhalten

$$\begin{aligned}\frac{\partial P(A)}{\partial p} &= \binom{10}{3} (3p^2(1-p)^7 - 7p^3(1-p)^6) \\ &= \binom{10}{3} p^2 \cdot (1-p)^6 \cdot (3(1-p) - 7p)\end{aligned}$$

- Nullsetzen der Ableitung liefert nun

$$\begin{aligned}\frac{\partial P(A)}{\partial p} = 0 &\Leftrightarrow 3(1-p) - 7p = 0 \\ &\Leftrightarrow 3 = 10p \\ &\Leftrightarrow p = 3/10\end{aligned}$$

Maximum Likelihood

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Sei nun im allgemeinen X eine Zufallsvariable mit **bekannter** Verteilung. Die Verteilung hänge von den Parametern θ ab.
- Angenommen wir haben n voneinander unabhängige Realisierungen (Beobachtungen) der Zufallsvariable, x_1, \dots, x_n gemacht. Dann ist

$$P(X|\theta) = \prod_{i=1}^n p_X(x_i|\theta)$$

die Wahrscheinlichkeit für diese Beobachtung.

- Bisher haben wir P stets als Funktion der Daten x_1, \dots, x_n betrachtet. Die Verteilung (und damit θ) war bekannt. Dies nannten wir *Wahrscheinlichkeit* der Daten.
- Hier sind aber die Daten bekannt (weil sie ja tatsächlich realisiert wurden), während die Parameter der Verteilung θ unbekannt sind.
- Dann nennen wir

$$\Lambda_X(\theta) = P(X|\theta) = \prod_{i=1}^n p_X(x_i|\theta)$$

likelihood der Parameter.

Maximum Likelihood Schätzung

Seien x_1, \dots, x_n unabhängige Realisierungen der Zufallsvariable X mit bekannter Verteilung aber unbekannter parametrisierung θ . Dann heisst

$$\theta^* = \arg \max_{\theta} \Lambda_X(\theta) = \arg \max_{\theta} \prod_{i=1}^n p_X(x_i|\theta)$$

Maximum Likelihood Schätzung von θ , gegeben X .

- Da das maximum dieses Produktterms oft schwer zu bestimmen ist, behilft man sich oft mit der so genannten *Log-Likelihood*

$$\lambda(\theta) = \ln \Lambda(\theta) = \sum_{i=1}^n \ln (p_X(x_i|\theta))$$

- Dies ist möglich da \ln streng monoton steigend ist.

Beispiel

- Aus Übung 6 ist die Exponentialverteilung bekannt. Ihre Dichte war

$$f_X(t) = \lambda \exp(-\lambda t)$$

für $t \geq 0$ und ihren Parameter $\lambda > 0$. Die Exponentialverteilung modelliert die Wartezeit auf das Auftreten eines Fehlers in einem System.

- Wir wollen die erwartete Lebensdauer einer Glühbirne abschätzen. Dazu sei eine Stichprobe von 3 Glühbirnen erhoben, wovon eine Birne 1000 Stunden, die zweite Birne 1200 Stunden und die dritte Birne 1500 Stunden gebrannt hat.
- Beschreibe X die Lebensdauer einer Glühbirne, dann ist X exponentialverteilt mit unbekanntem λ . Die Likelihood von λ gegeben obige Daten ist also

$$f_X(1000) \cdot f_X(1200) \cdot f_X(1700) = \lambda^3 \exp(-\lambda(1000 + 1200 + 1700))$$

Beispiel

- Wir suchen also das Maximum der Likelihoodfunktion

$$\Lambda(\lambda) = \lambda^3 \exp(-3900\lambda)$$

- Ableiten nach λ und nullsetzen liefert

$$\begin{aligned}\frac{\partial \Lambda}{\partial \lambda} &= 3\lambda^2 \exp(-3900\lambda) - 3900\lambda^3 \exp(-3900\lambda) = 0 \\ \Leftrightarrow (3 - 3900\lambda)\lambda^2 \exp(-3900\lambda) &= 0 \\ \Rightarrow \lambda = \frac{3}{3900} &\approx 0.00077\end{aligned}$$

- Für den Erwartungswert gilt (Vgl. Übung 6) demnach $E[X] = 1/\lambda = 3900/3 \approx 1300$. Eine Glühbirne wird also erwartungsgemäß etwa 1300 Stunden brennen.

ML-Schätzer für Exponentialverteilungen

Sei nun allgemein X exponentialverteilt mit unbekanntem λ und x_1, \dots, x_n eine Beobachtungsreihe. Dann ist

$$\lambda_{ML}^* = \frac{n}{\sum_{i=1}^n x_i}$$

die Maximum Likelihood Schätzung für λ gegeben x_1, \dots, x_n .

- **Beweis:** Die Likelihood Funktion ist

$$\Lambda(\lambda) = \lambda^n \exp \left(-\lambda \sum_{i=1}^n x_i \right)$$

- Ableiten und nullsetzen liefert dann die Behauptung.

ML-Schätzer für Normalverteilung

Es sei $X \sim \mathcal{N}(\mu, \sigma^2)$ normalverteilt mit unbekanntem μ und σ^2 , sowie x_1, \dots, x_n eine Beobachtungsreihe. Dann sind

$$\mu_{ML} = \frac{1}{n} \sum_{i=1}^n x_i$$

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

die Maximum Likelihood Schätzer für μ und σ^2 , gegeben x_1, \dots, x_n .

- Das bedeutet, die Maximum Likelihood Schätzung für den Erwartungswert μ und die Varianz σ^2 entspricht dem Mittelwert der Stichprobe bzw. der Varianz der Stichprobe.

Beweis

- Es sei $X \sim \mathcal{N}(\mu, \sigma^2)$ normalverteilt mit unbekannten Parametern μ und σ^2 . Zur Erinnerung: Die Dichtefunktion von X ist gegeben als

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Wir betrachten die Log-Likelihood von μ und σ^2 . Es ist

$$\lambda(\mu, \sigma^2) = \sum_{i=1}^n \ln f_X(x_i) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Beweis

$$\lambda(\mu, \sigma^2) = \sum_{i=1}^n \ln f_X(x_i) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

- Leiten wir die Log-Likelihood nach μ ab, erhalten wir

$$\begin{aligned} \frac{\partial \lambda}{\partial \mu} &= -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu) \cdot (-1) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \end{aligned}$$

- Nullsetzen liefert dann

$$\sum_{i=1}^n x_i - \sum_{i=1}^n \mu = 0 \Leftrightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Beweis

$$\lambda(\mu, \sigma^2) = \sum_{i=1}^n \ln f_X(x_i) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

- Leiten wir die Log-Likelihood nach σ^2 ab, erhalten wir

$$\frac{\partial \lambda}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

- Nullsetzen und Multiplikation mit $2\sigma^4$ liefert dann

$$-n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \Leftrightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Beispiel

- Wir erhalten aus einer normalverteilten Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ drei Beobachtungen $(x_1, x_2, x_3) = (10, 15, 17)$.
- Die Maximum Likelihood Schätzung für den Erwartungswert μ_{ML} beträgt

$$\mu_{ML} = \frac{10 + 15 + 17}{3} = 14$$

- Die Maximum Likelihood Schätzung für die Varianz σ_{ML}^2 beträgt

$$\sigma_{ML}^2 = \frac{(10 - 14)^2 + (15 - 14)^2 + (17 - 14)^2}{3} = 8.67$$

Maximum Likelihood

Wahrscheinlich

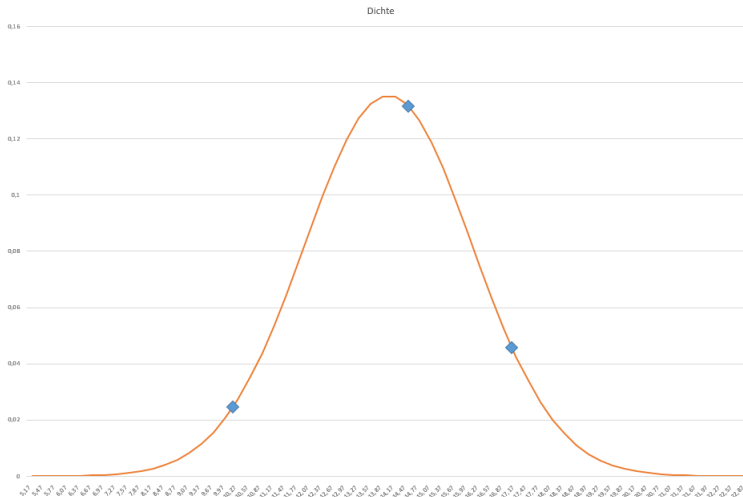
Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Beispiel



Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Parameterschätzung - Methode der kleinsten Quadrate

Einführendes Beispiel

- Mit unserem Fahrerassistenzsystem beobachten wir ein anderes Fahrzeug welches mit konstanter aber unbekannter Geschwindigkeit vor uns fährt.
- Wir machen drei Messungen des Abstandes d_1, d_2, d_3 zu den Zeitpunkten t_1, t_2, t_3 und erhalten

$t_i[s]$	0	1	3
$d_i[m]$	36	29	20

- Da das Fahrzeug mit konstanter Geschwindigkeit fährt, vermuten wir einen einfachen linearen Zusammenhang zwischen t_i und d_i , nämlich

$$d(t) = a + b \cdot t$$

- Wir suchen diejenigen Parameter a und b welche möglichst *gut* zu den Beobachtungen passen.

Einführendes Beispiel

- Als Gütekriterium wählen wir den quadratischen Fehler, also

$$\epsilon_i = (a + b \cdot t_i - d_i)^2$$

und suchen a und b so das die Summe dieser Fehler minimal wird, also die Fehlerfunktion

$$\mathcal{E}(a, b) = \sum_{i=0}^2 (a + b \cdot t_i - d_i)^2$$

minimiert.

- Die Ableitung nach a liefert

$$\frac{\partial \mathcal{E}}{\partial a} = 2 \sum_{i=0}^2 (a + b \cdot t_i - d_i) = 6a + 2b \sum_{i=0}^2 t_i - 2 \sum_{i=0}^2 d_i$$

Einführendes Beispiel

$$\mathcal{E}(a, b) = \sum_{i=0}^2 (a + b \cdot t_i - d_i)^2$$

$$\frac{\partial \mathcal{E}}{\partial a} = 2 \sum_{i=0}^2 (a + b \cdot t_i - d_i) = 6a + 2b \sum_{i=0}^2 t_i - 2 \sum_{i=0}^2 d_i$$

- Die Ableitung nach b liefert

$$\frac{\partial \mathcal{E}}{\partial b} = 2 \sum_{i=0}^2 (a + b \cdot t_i - d_i) t_i = 2a \sum_{i=0}^2 t_i + 2b \sum_{i=0}^2 t_i^2 - 2 \sum_{i=0}^2 d_i t_i$$

Einführendes Beispiel

- Einsetzen der obigen *Messungen* liefert dann dann Gleichungssystem

$$\begin{aligned}6a + 8b - 170 &= 0 \\8a + 20b - 178 &= 0\end{aligned}$$

- Auflösen liefert dann

$$a = 35.28m$$

und

$$b = -5.21m/s$$

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

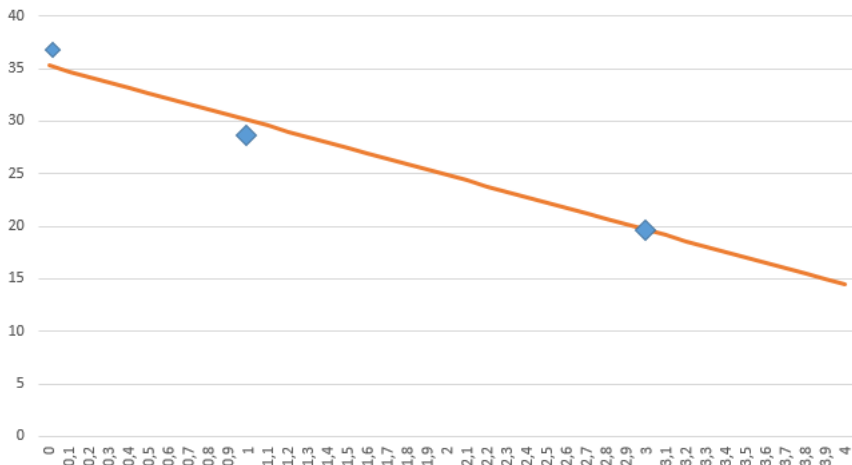
Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Einführendes Beispiel

Abstand über Zeit



Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Sei nun im allgemeinen eine Meßreihe $(x_1, y_1), \dots, (x_n, y_n)$ gegeben. Wir wollen den Zusammenhang zwischen den x_i und y_i über eine parametrisierbare Funktion f_θ modellieren, d.h. wir vermuten

$$y_i \approx f_\theta(x_i)$$

für eine geeignete Wahl von θ . θ ist im allgemeinen ein Vektor von k skalaren Parametern $\theta = (\theta_1, \dots, \theta_k)$.

- Die x_i heißen *unabhängige* Variablen während die y_i *abhängige* Variablen heißen (weil sie von x_i abhängen).
- Wir definieren das **Residuum**

$$r_i = y_i - f_\theta(x_i)$$

und die **Fehlerfunktion**

$$\mathcal{E}(\theta) = \sum_{i=1}^n r_i^2$$

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat.
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Wir wollen nun θ so bestimmen das \mathcal{E} minimal wird. Dazu bestimmen wir die partiellen Ableitungen von \mathcal{E} nach den Parametern $\theta_1, \dots, \theta_k$:

$$\frac{\partial \mathcal{E}}{\partial \theta_j} = 2 \sum_{i=1}^n r_i \frac{\partial r_i}{\partial \theta_j} = 0 \quad \forall j = 1, \dots, k$$

und wegen $r_i = y_i - f_{\theta}(x_i)$ erhalten wir

$$\frac{\partial \mathcal{E}}{\partial \theta_j} = -2 \sum_{i=1}^n r_i \frac{\partial f_{\theta}}{\partial \theta_j}$$

und falls θ die Fehlerfunktion minimiert gilt die *Least Squares Bedingung*

$$\forall j = 1, \dots, k : \quad \sum_{i=1}^n r_i \frac{\partial f_{\theta}}{\partial \theta_j} = 0$$

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat.
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Wir betrachten zunächst den Spezialfall linearer Modelle, d.h. das Modell ist linear in den Parametern $\theta_1, \dots, \theta_k$.
- Sei $f_\theta(x)$ linear in den Modelparametern $\theta_1, \dots, \theta_k$, d.h. es gilt

$$f_\theta(x) = \sum_{j=1}^k \theta_j \cdot \phi_j(x)$$

wobei die $\phi_j(x)$ nur von x , nicht mehr aber von den Parametern θ abhängen.

- Dann ist die Ableitung sehr einfach gegeben durch

$$\frac{\partial f_\theta}{\partial \theta_j} = \phi_j(x)$$

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Setzen wir in die allgemeine *Least Squares Bedingung* ein, so erhalten wir

$$\forall j = 1, \dots, k : \sum_{i=1}^n (y_i - f_{\theta}(x_i)) \cdot \phi_j(x_i) = 0$$

- Sei im folgenden $X_{ij} = \phi_j(x_i)$. Dann ist $f_{\theta}(x_i) = \sum_{m=1}^k \theta_m \cdot X_{im}$ und

$$\forall j = 1, \dots, k : \sum_{i=1}^n \left(y_i - \sum_{m=1}^k \theta_m \cdot X_{im} \right) \cdot X_{ij} = 0$$

- Umsortieren liefert dann

$$\forall j = 1, \dots, k : \sum_{i=1}^n X_{ij} \cdot y_i = \sum_{i=1}^n \sum_{m=1}^k X_{ij} X_{im} \theta_m$$

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Nun sei mit $X = (X_{ij})_{i=1,\dots,n;j=1,\dots,k}$ die Matrix aller X_{ij} bezeichnet, dann können wir auch schreiben

$$(X^T X) \cdot \theta = X^T y$$

mit $y = (y_1, \dots, y_n)^T$ dem Vektor der abhängigen Variablen.

- $(X^T X)$ ist stets eine quadratische Matrix und, i.d.R., invertierbar. Damit ist die gesuchte Lösung gegeben als

$$\theta = (X^T X)^{-1} X^T y$$

- Der Term $(X^T X)^{-1} X^T$ heisst *Pseudoinverse* von X .

Beispiel

- Wir betrachten wieder das Beispiel von gerade, also die $n = 3$ Messungen

$$\begin{array}{c|ccc} t_i[s] & 0 & 1 & 3 \\ \hline d_i[m] & 36 & 29 & 20 \end{array}$$

mit den $k = 2$ Parametern $\theta = (a, b)$ und dem Model

$$f_{\theta}(t) = a + bt$$

.

- Es ist $\phi_a(t) = 1$ und $\phi_b(t) = t$ und damit

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix} \quad y = \begin{pmatrix} 36 \\ 29 \\ 20 \end{pmatrix}$$

Beispiel

- Mit

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}$$

folgt also auch

$$(X^T X)^{-1} = \begin{pmatrix} 3 & 4 \\ 4 & 10 \end{pmatrix}^{-1} = \frac{1}{14} \begin{pmatrix} 10 & -4 \\ -4 & 3 \end{pmatrix}$$

und damit ist die Pseudoinverse von X

$$(X^T X)^{-1} X^T = \begin{pmatrix} 10 & 6 & -2 \\ -4 & -1 & 5 \end{pmatrix}$$

Die *Least Squares Schätzung* ist demnach (wie oben auch)

$$\theta_{LS} = (X^T X)^{-1} X^T \cdot y = \dots = (35.28; -5.21)^T$$

Beispiel

- Gehen wir nun davon aus das wir einen weiteren Meßpunkt $d_3 = 4$ zum Zeitpunkt $t_3 = 5$ erhalten haben. Unsere Beobachtungsreihe ist demnach

$t_i[s]$	0	1	3	5
$d_i[m]$	36	29	20	4

- Diesmal wollen wir als Model ein Polynom 2ten Grades annehmen, also $\theta = (a, b, c)$ mit

$$f_{\theta}(t) = a + bt + ct^2$$

- Es ist $\phi_a(x) = 1$, $\phi_b(t) = t$ und $\phi_c(t) = t^2$. Damit ist auch

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \end{pmatrix} \quad y = \begin{pmatrix} 36 \\ 29 \\ 20 \\ 4 \end{pmatrix}$$

Beispiel

- Die Pseudoinverse ist (selber nachrechnen!)

$$(X^T X)^{-1} X^T = \begin{pmatrix} 0.839 & 0.301 & -0.205 & 0.049 \\ -0.633 & 0.187 & 0.705 & -0.273 \\ 0.095 & -0.055 & -0.133 & 0.085 \end{pmatrix}$$

- Damit ist die *Least Squares Schätzung*

$$\theta_{LS} = (35.029; -4.357; -0.495)^T$$

- Dies entspricht dem Polynom

$$d(t) = 35.029 - 4.357t - 0.495t^2$$

Methode der kleinsten Quadrate

Wahrscheinlich

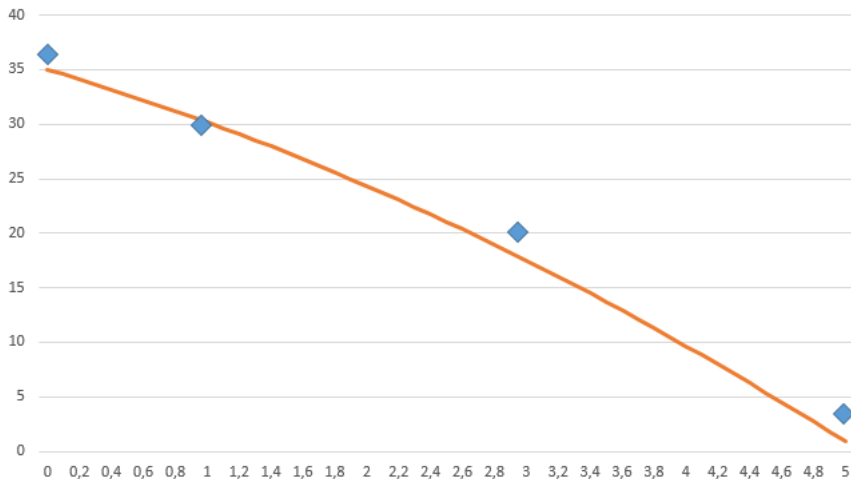
Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

Beispiel



Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat.
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Wir wollen nun zeigen, dass die Methode der kleinsten Quadrate und die Maximum-Likelihood Schätzung unter bestimmten Voraussetzungen identisch sind.
- Wir gehen davon aus das die Stichprobe y_1, \dots, y_n Realisationen von n Zufallsvariablen Y_i der Form

$$y_i = f_{\theta}(x_i) + \epsilon_i$$

mit bekanntem funktionalen Zusammenhang f_{θ} und normalverteiltem $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ ist.

- $f_{\theta}(x_i)$ beschreibt den von den Parametern θ abhängigen, funktionalen Zusammenhang, wobei die Beobachtung von diesem durch normalverteiltes Rauschen (Störeinfluß) überlagert wird.
- Dann gilt

$$E[Y_i] = E[f_{\theta}(X_i) + \epsilon_i] = E[f_{\theta}(X_i)]$$

■ Das Residuum

$$r_i = y_i - f_{\theta}(x_i) - \epsilon_i$$

ist somit ebenfalls Normalverteilt mit $r_i \sim \mathcal{N}(0, \sigma^2)$ und die Likelihood des Parametervektor θ ergibt sich zu

$$\Lambda(\theta) = P(Y|\theta) = \prod_{i=1}^n P(y_i|\theta) = \prod_{i=1}^n \mathcal{N}(r_i, 0, \sigma^2)$$

wobei hier

$$\mathcal{N}(r_i, 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r_i - 0)^2}{2\sigma^2}\right)$$

der Funktionswert der Normalverteilungsdichte an der Stelle r_i mit Erwartungswert 0 und Varianz σ^2 sein soll.

- Die Likelihood ist offensichtlich genau dann maximal wenn die individuellen Produktterme möglichst groß sind. Für die Normalverteilung ist dies der Fall wenn die Beobachtung (hier r_i) möglichst nahe am Erwartungswert liegt.

Methode der kleinsten Quadrate

Wahrscheinlich

Dr. rer. nat
Dennis
Müller

Motivation
und
Überblick

Maximum
Likelihood
Methode

Methode der
kleinsten
Quadrate

- Betrachten wir die Log-Likelihood und vernachlässigen additive Terme die nicht von θ abhängen, erhalten wir

$$\lambda(\theta) = \dots - \sum_{i=1}^n \frac{r_i^2}{2\sigma^2}$$

und damit wieder den selben *Least Squares* Schätzer wie oben.

- Dieser Zusammenhang gilt immer für den Fall einer Schätzung unter Einfluß von unabhängigen, normalverteiltem Meßrauschen.

Beispiel

- In einem Spiel kann man 1 Euro mit Wahrscheinlichkeit p verlieren, 1 Euro mit Wahrscheinlichkeit $1 - 2p$ gewinnen und mit Wahrscheinlichkeit p weder Geld verlieren noch gewinnen. Nach sechs spielen haben wir zwei mal verloren, 3 mal gewonnen und ein mal unentschieden gespielt. Wie groß ist p .

- **Maximum Likelihood Schätzung:**

Es ist

$$P = p^3 \cdot (1 - 2p)^3$$

und

$$\frac{\partial P}{\partial p} = 3p^2(1 - 2p)^3 - 6p^3(1 - 2p)^2 = 3p^2(1 - 2p)^2(1 - 4p)$$

und somit $p_{ML} = 1/4$.

Beispiel

- In einem Spiel kann man 1 Euro mit Wahrscheinlichkeit p verlieren, 1 Euro mit Wahrscheinlichkeit $1 - 2p$ gewinnen und mit Wahrscheinlichkeit p weder Geld verlieren noch gewinnen. Nach sechs spielen haben wir zwei mal verloren, 3 mal gewonnen und ein mal unentschieden gespielt. Wie groß ist p .
- **Methode der kleinsten Quadrate:**
Es sei X der Gewinn des Spiels. Dann ist

$$E[X] = -1 \cdot p + 0 \cdot +1 \cdot (1 - 2p) = 1 - 3p$$

Wir minimieren die quadratische Abweichung der Beobachtungen zum Erwartungswert, also

$$2(-1 - (1 - 3p))^2 + 3(1 - (1 - 3p))^2 + 1(0 - (1 - 3p))^2 = 9 - 30p + 54p^2$$

und damit $p_{LS} = 5/18$.

Was Sie heute gelernt haben

- Sie verstehen die Maximum Likelihood Methode als Schätzmethode, welche die Parameter so bestimmt, dass die Wahrscheinlichkeit für die beobachteten Daten maximal wird.
- Sie verstehen die Methode der kleinsten Quadrate als Schätzmethode, welche die Parameter so bestimmt, dass die quadratischen Abweichungen zwischen Beobachtungen und erwarteten Beobachtungen minimiert wird.
- Sie wissen, dass dies das selbe ist, falls die Beobachtungen durch unabhängiges, normalverteiltes Rauschen überlagert werden.