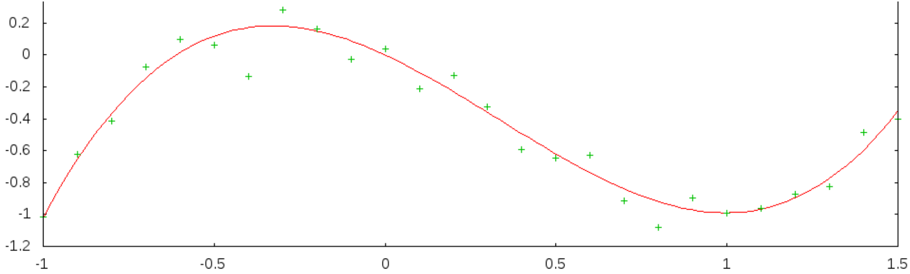
Anpassungsrechnungen mit kleinsten Quadraten und Maximum Likelihood



Gliederung

Das Problem

0.5159	6.1404	0.0409	0.004181
1.2335	0.3405	0.0819	0.004434
2.0768	0.0963	0.1228	0.004827
3.0872	0.0525	0.1637	0.005328
4.0816	0.0294	0.2046	0.005910
4.8890	0.0188	0.2456	0.006552
5.5414	0.0161	0.2865	0.007237
6.0929	0.0123	0.3274	0.007955
6.5342	0.0123	0.3684	0.008698
6.8709	0.0122	0.4093	0.009459
7.1006	0.0086	0.4502	0.010234
7.2838	0.0085	0.4912	0.011021
7.3960	0.0080	0.5321	0.011817
7.4743	0.0069	0.5730	0.012620
7.4906	0.0067	0.6139	0.013430
7.4860	0.0064	0.6549	0.014244
7.4650	0.0080	0.6958	0.015063
7.4396	0.0063	0.7367	0.015885

- Man hat viele Daten
- Die Theorie hat wenige Parameter
- Parameter müssen so angepasst werden, dass sie gut zu den Daten passen

Die Lösung

Anpassungsrechnung (auch Parameterschätzung, Parameteranpassung, Ausgleichungsrechnung, Fitting) mit

- Maximum-Likelihood-Methode
- Methode der kleinsten Quadrate

Wahrscheinlichkeitsverteilung

Eine Wahrscheinlichkeitsverteilung kann

- diskret (Bsp. Würfel, Spin) oder
- kontinuierlich

sein.

Wahrscheinlichkeitsverteilung

kontinuierliche Wahrscheinlichkeit

Definition

Wahrscheilichkeitsdichtefunktion (probability density function, pdf)

$$P\left(a < x < b\right) = \int_a^b f\left(x\right) dx$$

mit

$$\int_{-\infty}^{\infty} f\left(x\right) dx = 1$$

Erwartungswert und Varianz

Definition

Erwartungswert einer Funktion $h(x)$:

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) f(x) \, dx$$

Erwartungswert und Varianz

Definition

Erwartungswert einer Funktion $h(x)$:

$$E[h(x)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx$$

Definition

Mittelwert=Erwartungswert von x

$$\langle x \rangle = E[x] = \int_{-\infty}^{\infty} x f(x)\, dx$$

Erwartungswert und Varianz

Definition

Varianz von x:

$$V[x] = E[(x - \langle x \rangle)^2] = E[x^2] - E[x]^2$$

Erwartungswert und Varianz

Definition

Varianz von x:

$$V[x] = E[(x - \langle x \rangle)^2] = E[x^2] - E[x]^2$$

Definition

Kovarianz von x und y:

$$cov[x, y] = E[(x - \langle x \rangle)(y - \langle y \rangle)] = E[x \cdot y] - E[x]E[y]$$

Erwartungswert und Varianz

Definition

Kovarianzmatrix:

$$\mathbf{V}[\mathbf{x}] = E[(\mathbf{x} - \langle\mathbf{x}\rangle)(\mathbf{x} - \langle\mathbf{x}\rangle)^T]$$

$$\mathbf{V} = \begin{pmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \dots & cov(x_1, x_n) \\ cov(x_2, x_1) & cov(x_2, x_2) & \dots & cov(x_2, x_n) \\ \vdots & \vdots & & \vdots \\ cov(x_n, x_1) & cov(x_n, x_2) & \dots & cov(x_n, x_n) \end{pmatrix}$$

Zentraler Grenzwertsatz

S sei die Summe aus N unabhängigen Zufallsvariablen mit Mittelwert a_i und Varianz V_i. Dann gilt:

- $E[S] = \sum_{i=1}^{N} a_i$
- $V[S] = \sum_{i=1}^{N} V_i$
- Für $N \to \infty$ ist die pdf von S eine Normalverteilung

Eigenschaften von Schätzern

Ein Schätzer sollte folgende Eigenschaften haben:

- Konsistenz
- Erwartungstreue
- Effizienz
- Robustheit

Effizienz und Robustheit stehen oft in Widerspruch zueinander.

- Benötigte Rechenleistung sollte gering sein

Eigenschaften von Schätzern

Ein Schätzer sollte folgende Eigenschaften haben:

- Konsistenz: $\lim_{N \to \infty} \hat{a} = a_0$
- Erwartungstreue
- Effizienz
- Robustheit

Effizienz und Robustheit stehen oft in Widerspruch zueinander.

- Benötigte Rechenleistung sollte gering sein

Eigenschaften von Schätzern

Ein Schätzer sollte folgende Eigenschaften haben:

- Konsistenz: $\lim_{N \to \infty} \hat{a} = a_0$
- Erwartungstreue: $E[\hat{a}] = a_0$
- Effizienz
- Robustheit

Effizienz und Robustheit stehen oft in Widerspruch zueinander.

- Benötigte Rechenleistung sollte gering sein

Eigenschaften von Schätzern

Ein Schätzer sollte folgende Eigenschaften haben:

- Konsistenz: $\lim_{N\to\infty} \hat{a} = a_0$
- Erwartungstreue: $E[\hat{a}] = a_0$
- Effizienz: $V[\hat{a}]$ sollte minimal sein
- Robustheit

Effizienz und Robustheit stehen oft in Widerspruch zueinander.

- Benötigte Rechenleistung sollte gering sein

Eigenschaften von Schätzern

Ein Schätzer sollte folgende Eigenschaften haben:

- Konsistenz: $\lim\limits_{N \to \infty} \hat{a} = a_0$
- Erwartungstreue: $E[\hat{a}] = a_0$
- Effizienz: $V[\hat{a}]$ sollte minimal sein
- Robustheit: Kein Einfluss von falschen Daten/Voraussetzungen

Effizienz und Robustheit stehen oft in Widerspruch zueinander.

- Benötigte Rechenleistung sollte gering sein

Maximum-Likelihood-Methode

Definition

Likelihood-Funktion

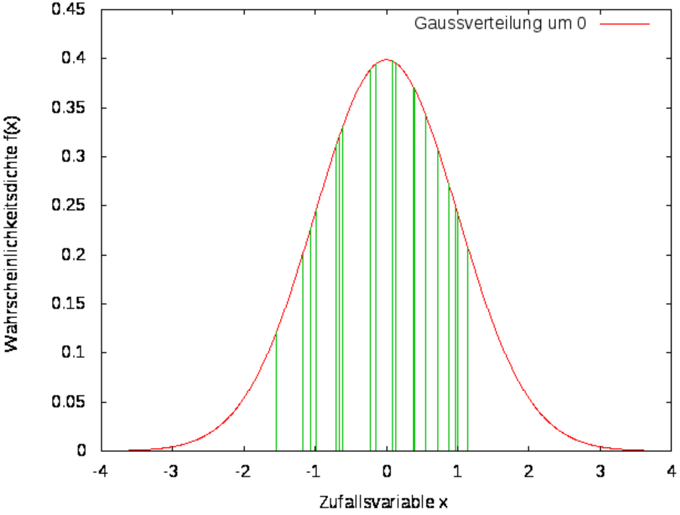
$$L(\mathbf{a}) = \prod_i f(x_i, \mathbf{a})$$

beschreibt die Wahrscheinlichkeit die Messwerte $\{x_i\}$ zu erhalten.

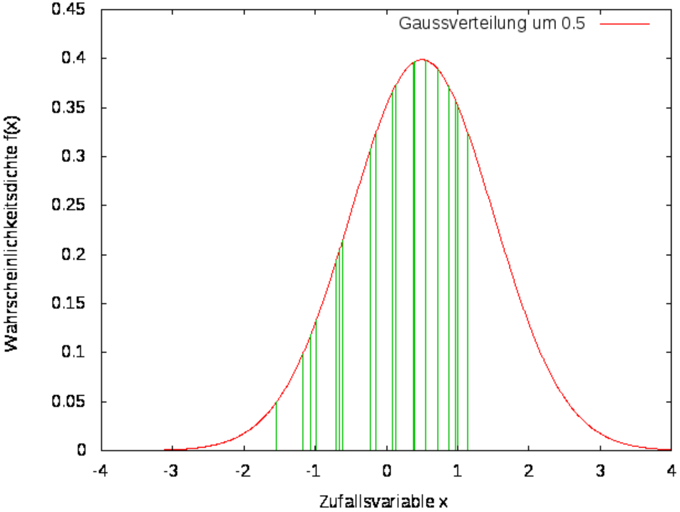
- Maximum-Likelihood-Methode: Maximierung der Likelihood-Funktion
- Praktisch meist Minimierung der negative Log-Likelihood-Funktion:

$$F(\mathbf{a}) = -\ln L(\mathbf{a}) = -\sum_i \ln f(x_i, \mathbf{a})$$

Maximum-Likelihood-Methode



Maximum-Likelihood-Methode



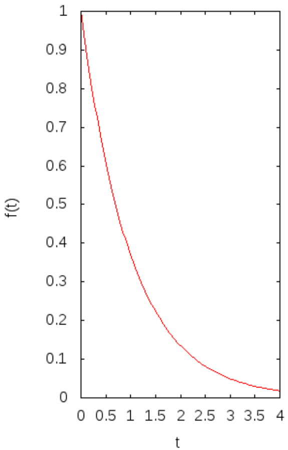
Einfaches Beispiel
Lebensdauer

$$f(t,\gamma) = \gamma e^{-\gamma t} \qquad \text{mit } \gamma = \frac{1}{\tau}$$

$$F(\gamma) = -N \cdot \ln(\gamma) + \sum_{i=1}^{N} \gamma t_i$$

$$\frac{d}{d\gamma} F(\gamma) = -\frac{N}{\gamma} + \sum_{i=1}^{N} t_i \stackrel{!}{=} 0$$

$$\hat{\tau} = \frac{1}{\hat{\gamma}} = \sum_{i=1}^{N} \frac{t_i}{N}$$



Einfaches Beispiel

Lebensdauer

Angenommen es kann nur im Zeitintervall $0 < t < t_{max}$ gemessen werden.

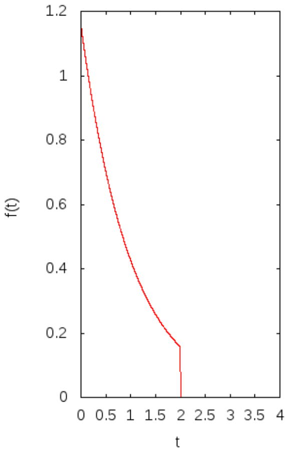
$$f(t, \gamma) = \frac{\gamma e^{-\gamma t}}{1 - e^{-\gamma t_{max}}}$$

$$F(\gamma) = N \left[\ln \left(1 - e^{-\gamma t_{max}} \right) - \ln(\gamma) \right] + \sum_{i=1}^{N} \gamma t_i$$

$$\vdots$$

$$\hat{\tau} = \sum_{i=1}^{N} \frac{t_i}{N} + \frac{t_{max} \exp(-t_{max}/\hat{\tau})}{1 - \exp(-t_{max}/\hat{\tau})}$$

Dies ist nur numerisch auswertbar.

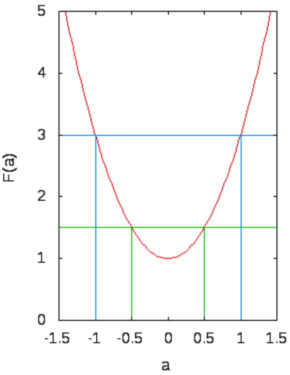


Fehlerberechnung

Parabolischer Fall

Bei großen
Proben ist die Fehlerzuordnung einfach:

- Nach dem Zentralen Grenzwertsatz geht die Likelihood-Funktion gegen eine Gauß-Funktion, bei großen Stichproben.
- Entwicklung:
  $F(a) = F(\hat{a}) + \frac{1}{2} \cdot \frac{d^2 F}{da^2} \cdot (\hat{a} - a)^2 + \mathcal{O}(a^3)$
- Vergleich
  mit Gauß: $-\ln(\mathrm{Gauß}) = const + \frac{1}{2} \frac{(\hat{a}-a)^2}{\sigma^2}$
- $\Rightarrow F(\hat{a} \pm r \cdot \sigma) = F(\hat{a}) + \frac{r^2}{2}$

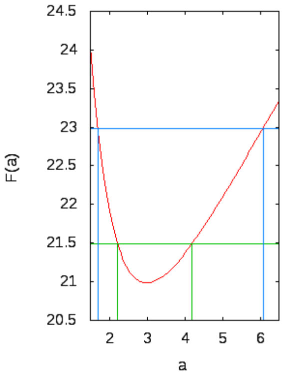


Fehlerberechnung

Allgemeiner Fall

Es können asymmetrische Konfidenzgrenzen definiert werden:

- $F(\hat{a} + r \cdot \sigma_+) = F(\hat{a}) + \frac{r^2}{2}$
- $F(\hat{a} - r \cdot \sigma_-) = F(\hat{a}) + \frac{r^2}{2}$



Eigenschaften der Maximum-Likelihood-Methode

- Schätzer ist invariant unter Parameter-Transformation: $\hat{g}(a) = g(\hat{a})$
- Normalerweise konsistent
- Nicht immer erwartungstreu
- Erwartungstreu für $N \to \infty$
- Effizienteste Methode
- Nicht robust, da Wahrscheinlichkeitsdichte exakt bekannt sein muss
- Kann hohe Rechenleistung erfordern

Methode der kleinsten Quadrate

Karlsruher Institut für Technologie

Definition

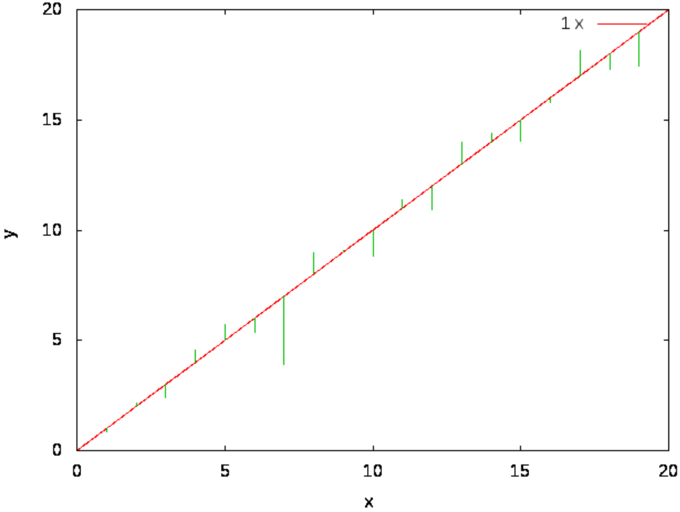
$$S = \sum_{i=1}^{N} \frac{(y_i - a_i)^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{\Delta y_i^2}{\sigma_i^2}$$

- N Messwerte y_i mit Varianz σ_i^2
- a_i aus Theorie
- Δy_i heißen Residuen
- Methode der kleinsten Quadrate: Minimiere S
- Allgemeiner Fall mit Kovarianzmatrix $\mathbf{V}$ und $\Delta\mathbf{y} = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_N \end{pmatrix}$

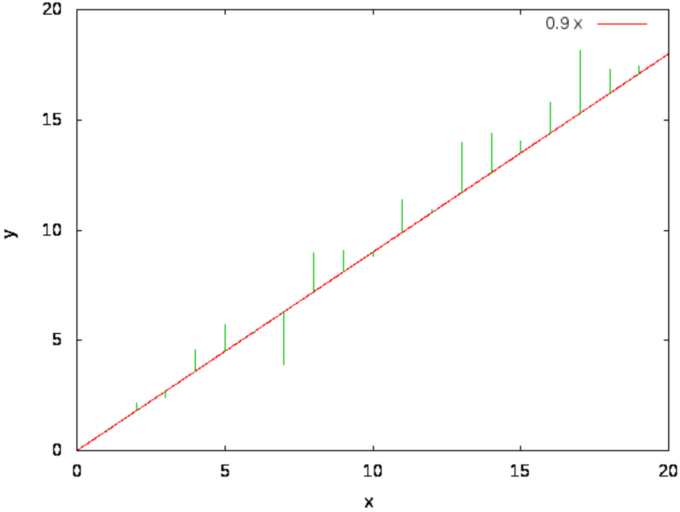
$$S = \Delta\mathbf{y}^T \mathbf{V}^{-1} \Delta\mathbf{y}$$

Methode der kleinsten Quadrate

Karlsruher Institut für Technologie



Methode der kleinsten Quadrate



Zusammenhang von Maximum-Likelihood und Kleinste Quadrate

negative log-Likelihood für gaußsche Wahrscheinlichkeitsdichte.

$$F(a) = -\sum_i \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\frac{(x_i-a)^2}{\sigma^2}}\right)$$

$$= const + \frac{1}{2}\sum_i \frac{(x_i - a)^2}{\sigma^2}$$

$$= const + \frac{1}{2} S(a)$$

Lineare Kleinste Quadrate

Karlsruher Institut für Technologie

- Lineare Theorie $t(x|\mathbf{a}) = \sum_i a_i t_i(x)$
- Zur Vereinfachung gleiche Varianzen
- $S = \frac{1}{\sigma^2} \sum_i (y_i - t(x_i))^2$
- $\frac{\partial S}{\partial a_j} = -\frac{2}{\sigma^2} \sum_i t_j(x_i)(y_i - t(x_i)) \stackrel{!}{=} 0$
- Normalengleichung: $\sum_i t_j(x_i) \sum_k \hat{a}_k t_k(x_i) = \sum_i y_i t_j(x_i)$

Lineare Kleinste Quadrate

- Matrixschreibweise

$$\mathbf{A} = \begin{pmatrix} t_1(x_1) & t_2(x_1) & \dots & t_p(x_1) \\ t_1(x_2) & t_2(x_2) & \dots & t_p(x_2) \\ \vdots & & & \vdots \\ t_1(x_n) & t_2(x_n) & \dots & t_p(x_n) \end{pmatrix}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- Normalengleichung: $(\mathbf{A}^T\mathbf{A})\hat{\mathbf{a}} = \mathbf{A}^T\mathbf{y}$

$$\hat{\mathbf{a}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$$

- Allgemeine Lösung mit Gewichtsmatrix $\mathbf{W} = \mathbf{V}^{-1}$:

$$\hat{\mathbf{a}} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y} = \mathbf{B}\mathbf{y}$$

- Fehlerfortpflanzung:

$$\mathbf{V_a} = \mathbf{B}\mathbf{y}\mathbf{B}^T$$

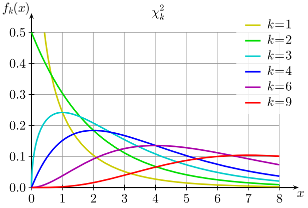
$$\mathbf{V_a} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{V}\left((\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\right)^T$$

$$\mathbf{V_a} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}A^T\mathbf{W}\mathbf{A}(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$$

$$\mathbf{V_a} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$$

$\mathcal{X}^2$-Test

- $\mathcal{X}_k^2$-Verteilung gibt die Wahrscheinlichkeitsdichte für die Summe der Quadrate von k standard-normalverteilten Zufallsvariablen (k Freiheitsgrade) an.
- Für gaußsche Fehler folgt S einer $\mathcal{X}^2$ Verteilung mit $n - p$ Freiheitsgraden.
- Wahrscheinlichkeit den Wert S oder größeren zu bekommen: $\int_S^\infty = f(\mathcal{X}^2) d\mathcal{X}^2$
- Wenn diese Wahrscheinlichkeit zu klein ist (z.B 1%), dann wird die Theorie verworfen.

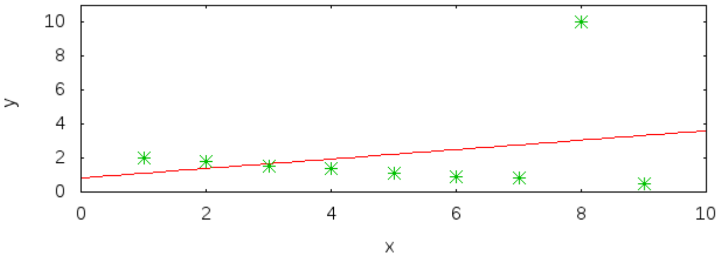


Eigenschaften von linearen Kleinsten Quadraten

- Es gibt analytische Lösung
- Konsistent und erwartungstreu, wenn Daten unverzerrt sind
- Effizienteste lineare erwartungsteue Methode
- Nicht Robust, wegen Außreißern in den Daten
- Benötigte Rechenleistung eher gering
- Bei gaußschen Fehlern Fit-Qualität mit $\mathcal{X}^2$

Eigenschaften von linearen Kleinsten Quadraten

- Es gibt analytische Lösung
- Konsistent und erwartungstreu, wenn Daten unverzerrt sind
- Effizienteste lineare erwartungsteue Methode
- Nicht Robust, wegen Außreißern in den Daten
- Benötigte Rechenleistung eher gering
- Bei gaußschen Fehlern Fit-Qualität mit $\mathcal{X}^2$



Zusammenfassung

	Maximum-Likelihood	Kleinste Quadrate
Voraussetzung	pdf exakt bekannt	Mittelwerte und Varianzen
Konsistent	Ja	Ja
Erwartungstreu	Nur asymptotisch	Im linearen Fall
Effizient	maximal	maximal
Robust	Nein (pdf muss exakt bekannt sein)	Nein (Ausreißer)
Rechenaufwand	kann sehr hoch werden	im linearen Fall gering
Fit-Qualität	nein	bei gaußschen Fehlern

Gleich bei gaußschen Fehlern

Literatur

- Volker Blobel, Erich Lohrmann: *Statistische und numerische Methoden der Datenanalyse*
- Gerhard Bohm, Günter Zech: *Einführung in Statistik und Messwertanalyse für Physiker*
  `http://www-library.desy.de/preparch/books/vstatmp.pdf`
- Roger Barlow: *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*