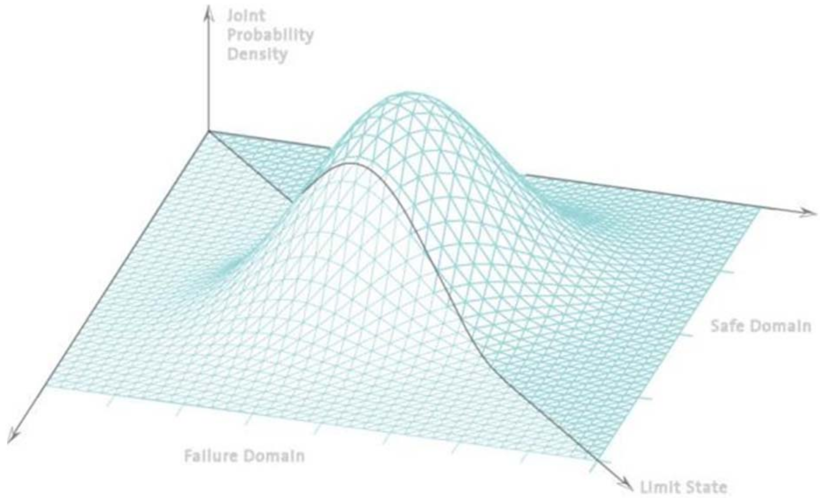


Statistik und Wahrscheinlichkeitsrechnung

9. Vorlesung

Jochen Köhler

Testatprüfung am Donnerstag 5.Mai

Wann?

Donnerstag, 5. Mai, 8:00 Uhr

Dauer der Prüfung: 60 min

Wo?

Die Raumaufteilung wird noch auf unserer Homepage veröffentlicht:

www.ibk.ethz.ch/fa/education/ss_statistics

Testatprüfung am Donnerstag 5.Mai

Inhalt

60 min. Multiple Choice

Gesamter Stoff bis einschliesslich Vorlesung 9 und Übung 9.

Hilfsmittel

Eine DIN A4 Seite doppelseitig Formelsammlung erlaubt.

Keine weiteren Hilfsmittel erlaubt!

Inhalte der heutigen Vorlesung

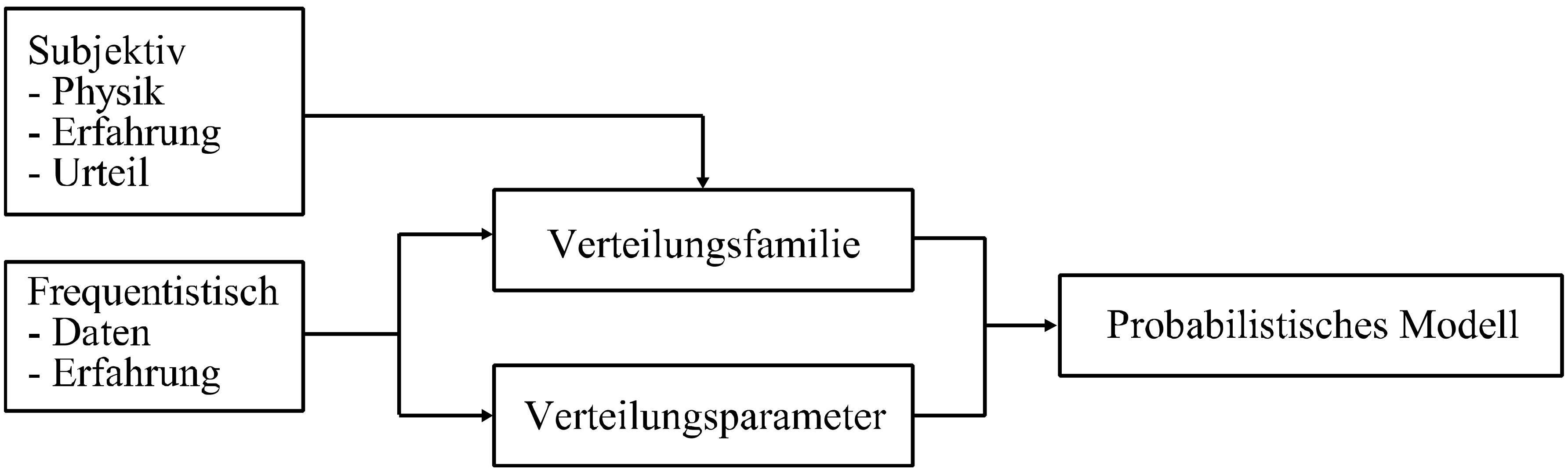
- Bayes‘sche Methoden der Parameterschätzung
  - Mathematische Formulierung
  - Konjugierte *a priori* Verteilungen

- Bayes‘sche Regressionsanalyse
  - Formulierung eines Regressionsmodells
  - Methode der kleinsten Quadrate
  - Aktualisierung der Regressionskoeffizienten

Modellierung von Unsicherheiten - Übersicht

Unterschiedliche Typen an Informationen werden zur Bildung von Ingenieurmodellen verwendet

- Subjektive Information
- Frequentistische Information



Bayes'sche Methoden der Parameterschätzung

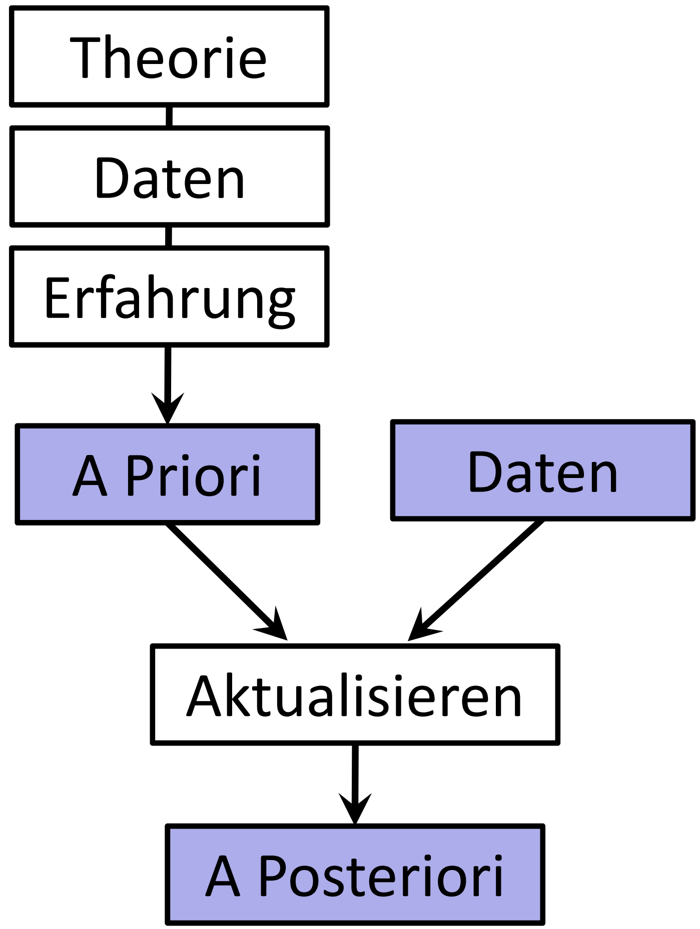
Bayes'sche Methoden der Parameter-schätzung ermöglichen die gleichzeitige Berücksichtigung von

a priori (subjektiver und/oder frequentistischer) Information

und

zusätzlicher frequentistischer Information (Versuchsergebnisse oder Beobachtungen)

So kann ein probabilistisches Modell aktualisiert werden → *a posteriori* Modell



Bayes'sche Methoden der Parameterschätzung

Wir betrachten die Zufallsvariable X mit der Wahrscheinlichkeitsdichtefunktion

$$f_X(x, \boldsymbol{\theta})$$

Die Parameter $\boldsymbol{\theta}$ der Dichtefunktion werden als unsicher angenommen mit der *a priori* Wahrscheinlichkeitsdichte

$$f'_\Theta(\boldsymbol{\theta})$$

n Versuche / Beobachtungen werden gemacht:

$$\hat{\mathbf{x}} = \left(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n\right)^T$$

Bayes'sche Methoden der Parameterschätzung

Die Likelihood der Beobachtungen berechnet sich zu:

$$L(\boldsymbol{\theta}|\hat{\mathbf{x}}) = \prod_{i=1}^{n} f_X(\hat{x}_i|\boldsymbol{\theta})$$

Mit dem Satz von Bayes erhalten wir die *a posteriori* Wahrscheinlichkeitsdichtefunktion für die unsicheren Parameter $\boldsymbol{\Theta}$:

$$f''_{\Theta}(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

Nun können wir die Dichtefunktion für X aktualisieren. So erhalten wir die (a posteriori) *prädiktive* Dichtefunktion:

$$f_X(x|\mathbf{x}) = \int f_X(x|\boldsymbol{\theta})\, f''_{\Theta}(\boldsymbol{\theta}|\hat{\mathbf{x}})\, d\boldsymbol{\theta}$$

Bayes‘sche Methoden der Parameterschätzung

Es gibt eine grosse Zahl von natürlich konjugierten Verteilungen für

- die a priori
- die a posteriori

Wahrscheinlichkeitsdichtefunktionen.

Bei natürlich konjugierten Verteilungen gehören sowohl die a priori als auch die a posteriori Wahrscheinlichkeitsdichtefunktion zur selben Verteilungsfamilie.

Bayes'sche Methoden der Parameterschätzung

Illustration: Aktualisieren einer Dichtefunktion

$$f''_{\Theta}(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

Die Parameter $\boldsymbol{\Theta}$ sind unsicher, z.B.

- Mittelwert der Fliessgrenze von Stahl

- Standardabweichung einer Schneelast

→ statistische Unsicherheit

Bayes‘sche Methoden der Parameterschätzung

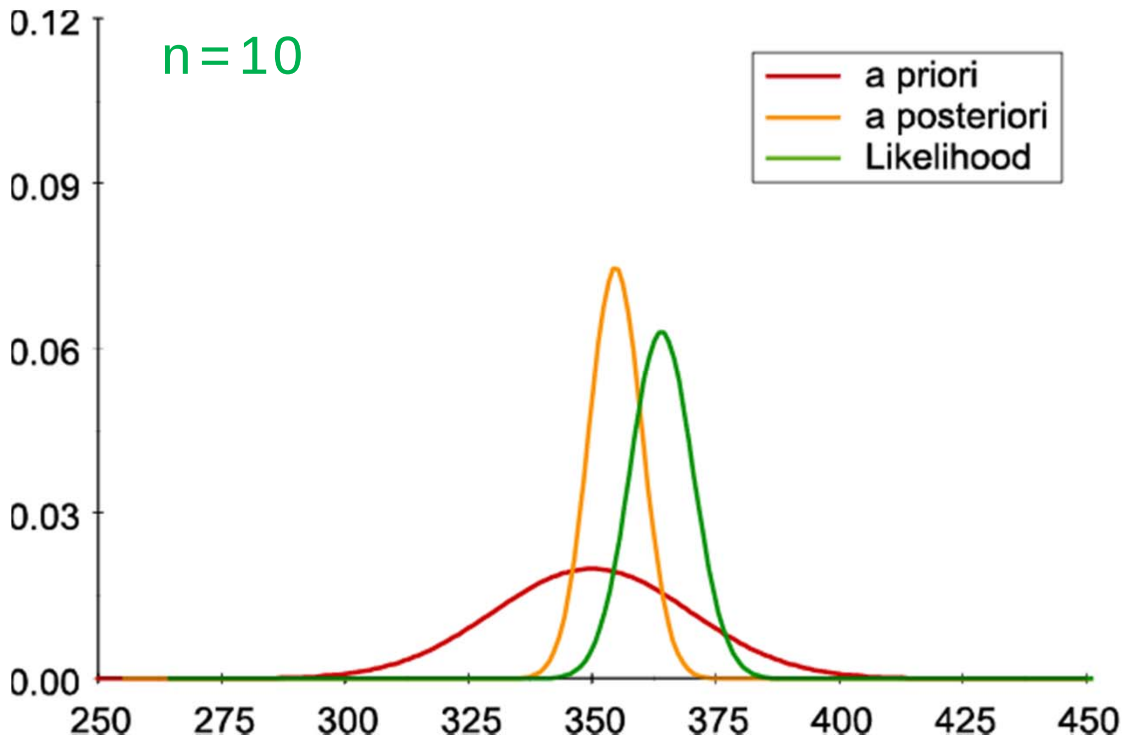
Illustration: Aktualisieren einer Dichtefunktion

Likelihood

a priori

$$f''_{\Theta}(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

a posteriori



Bayes‘sche Methoden der Parameterschätzung

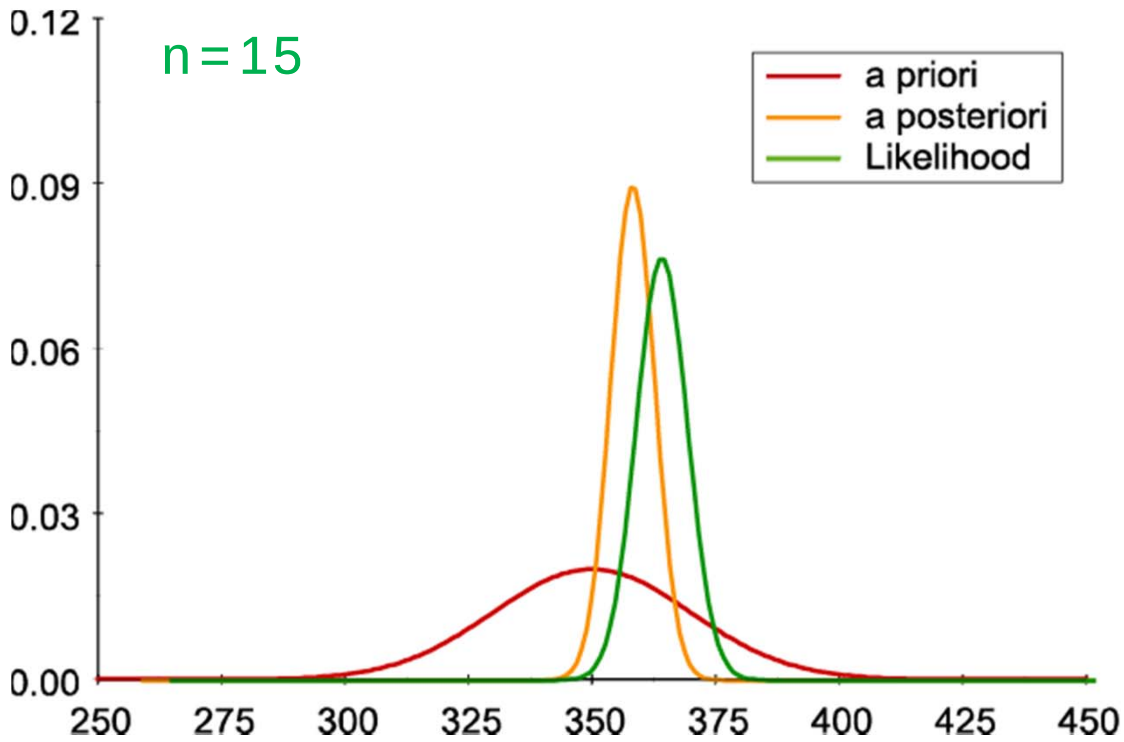
Illustration: Aktualisieren einer Dichtefunktion

Likelihood

a priori

$$f''_{\Theta}(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

a posteriori



Je grösser die Anzahl Daten, desto stärker ist der Einfluss der Likelihood

Bayes'sche Methoden der Parameterschätzung

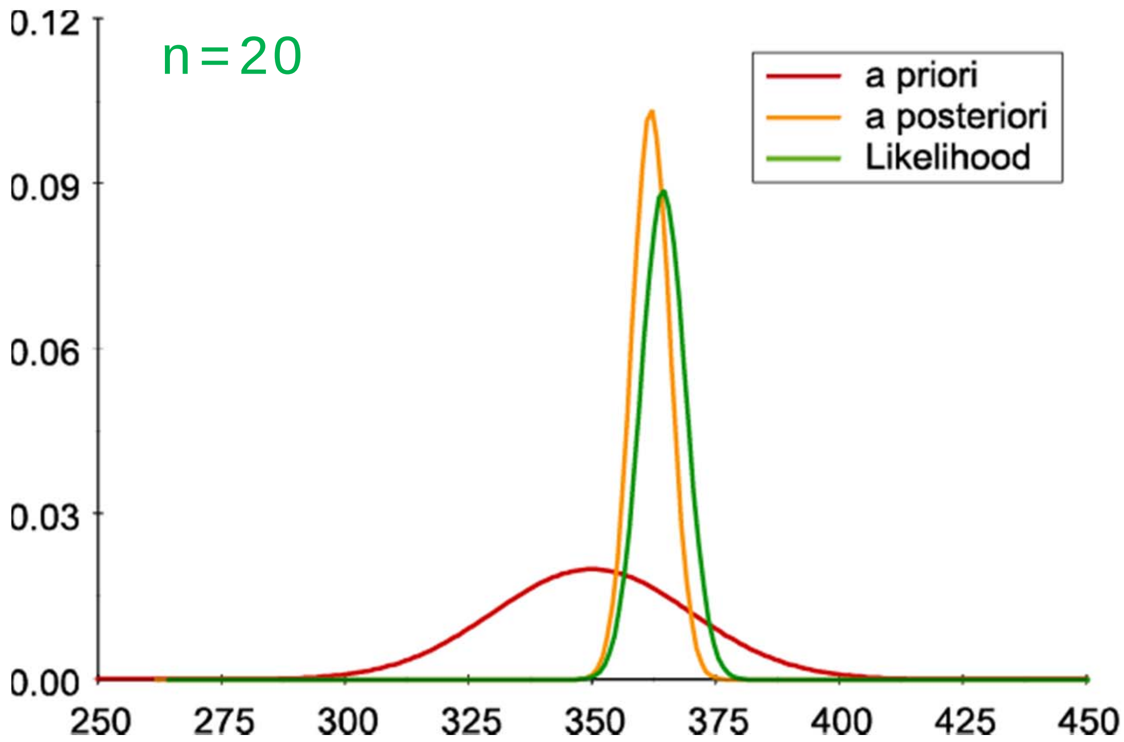
Illustration: Aktualisieren einer Dichtefunktion

Likelihood

a priori

$$f''_{\Theta}(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\hat{\mathbf{x}})\, f'_{\Theta}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

a posteriori



Je grösser die Anzahl Daten, desto stärker ist der Einfluss der Likelihood

Bayes'sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl

Wir nehmen an, die Fliessspannung f_y eines Baustahls in MPa sei normalverteilt mit Erwartungswert μ_{f_y} und Standardabweichung σ_{f_y}:

$$f_y \sim N(\mu_{f_y};\sigma_{f_y})$$

Die Standardabweichung sei bekannt, der Erwartungswert unsicher:

$$\sigma_{f_y} = 17.5\,MPa \qquad \mu_{f_y} \sim N(\mu';\sigma') \quad \text{mit} \quad \begin{aligned} \mu' &= 350MPa \\ \sigma' &= 10MPa \end{aligned}$$

Zusätzlich wurden 5 Zugversuche durchgeführt und die folgenden Fliessspannungen gemessen:

$$\hat{\mathbf{f}}_y = (365, 347, 354, 362, 348)$$

Bayes‘sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl

Bei einer normalverteilten Zufallsvariable mit bekannter Standard-abweichung und unbekanntem, normalverteiltem Erwartungswert ist die *a posteriori* Verteilung des Erwartungswerts ebenfalls normalverteilt:

$$f''(\mu_{f_y} | \hat{\mathbf{f}}_y) = \varphi(\mu_{f_y}) = \frac{1}{\sqrt{2\pi}\sigma''} \exp\left(-\frac{1}{2}\left(\frac{\mu_{f_y} - \mu''}{\sigma''}\right)^2 \right)$$

Die Parameter der a posteriori Verteilung sind:

$$n' = \frac{\sigma_{f_y}^2}{\sigma'^2} \qquad \mu'' = \frac{\frac{\mu'}{n} + \frac{\bar{f}_y}{n'}}{\frac{1}{n'} + \frac{1}{n}} \qquad \sigma'' = \sqrt{\frac{\frac{\sigma_{f_y}^2}{n'} \cdot \frac{\sigma'^2}{n}}{\frac{\sigma'^2}{n'} + \frac{\sigma'^2}{n}}}$$

Bayes‘sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl

Bei einer normalverteilten Zufallsvariable mit bekannter Standard-abweichung und unbekanntem, normalverteiltem Erwartungswert ist die *a posteriori* Verteilung des Erwartungswerts ebenfalls normalverteilt:

$$f''(\mu_{f_y} | \hat{\mathbf{f}}_y) = \varphi(\mu_{f_y}) = \frac{1}{\sqrt{2\pi}\sigma''} \exp\left(-\frac{1}{2}\left(\frac{\mu_{f_y} - \mu''}{\sigma''} \right)^2 \right)$$

Im betrachteten Beispiel:

$$n' = 3.06 \qquad \mu'' = 353.22\ MPa \qquad \sigma'' = 6.16\ MPa$$

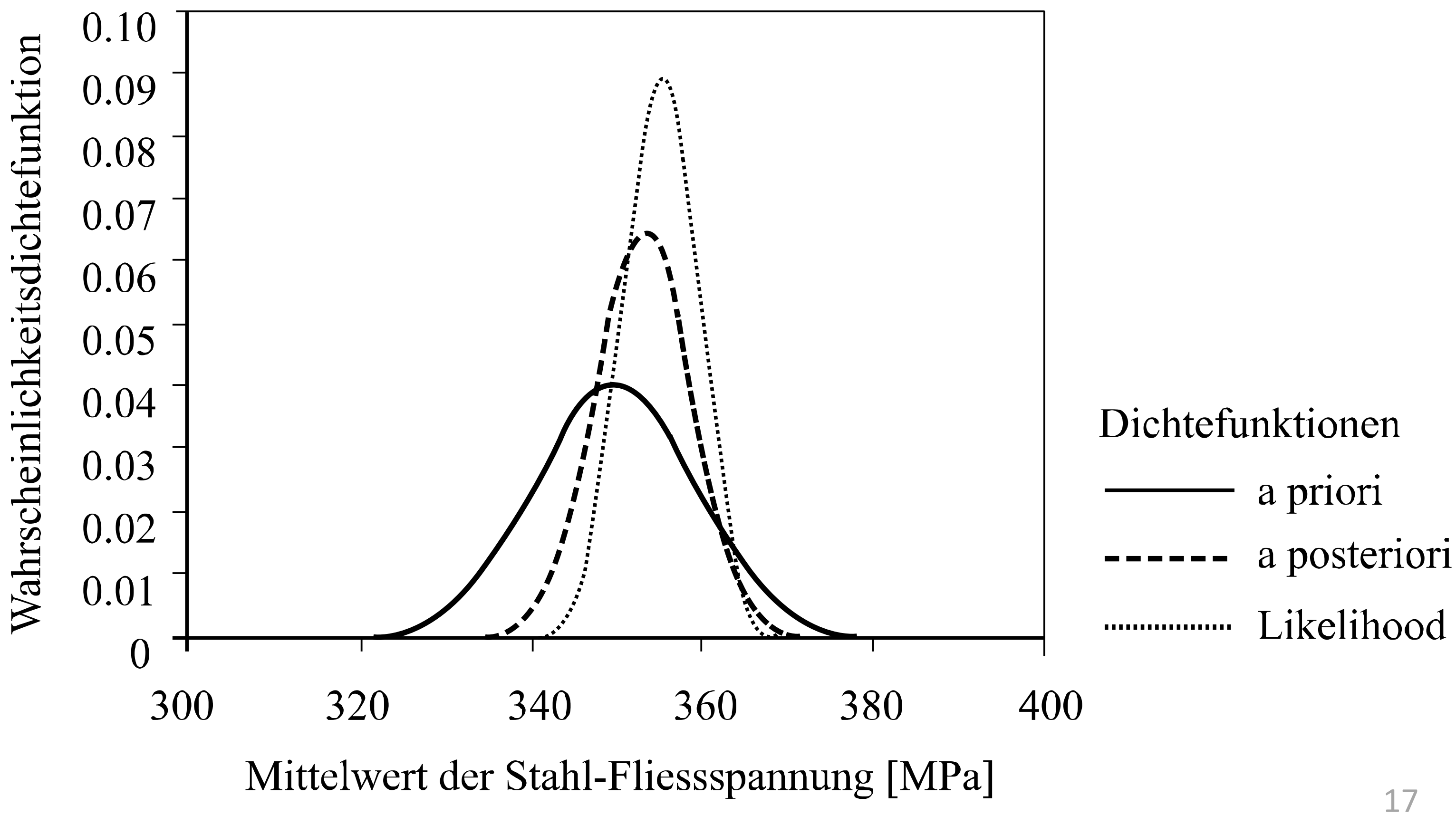
Bayes'sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl

Im Beispiel hat die Likelihood einen relativ starken Einfluss

$$n' = 3.06$$

$$n = 5$$



Bayes'sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl

Schliesslich können wir auch die (a posteriori) *prädiktive* Wahrscheinlichkeitsdichtefunktion für die Stahl-Fliessspannung bestimmen:

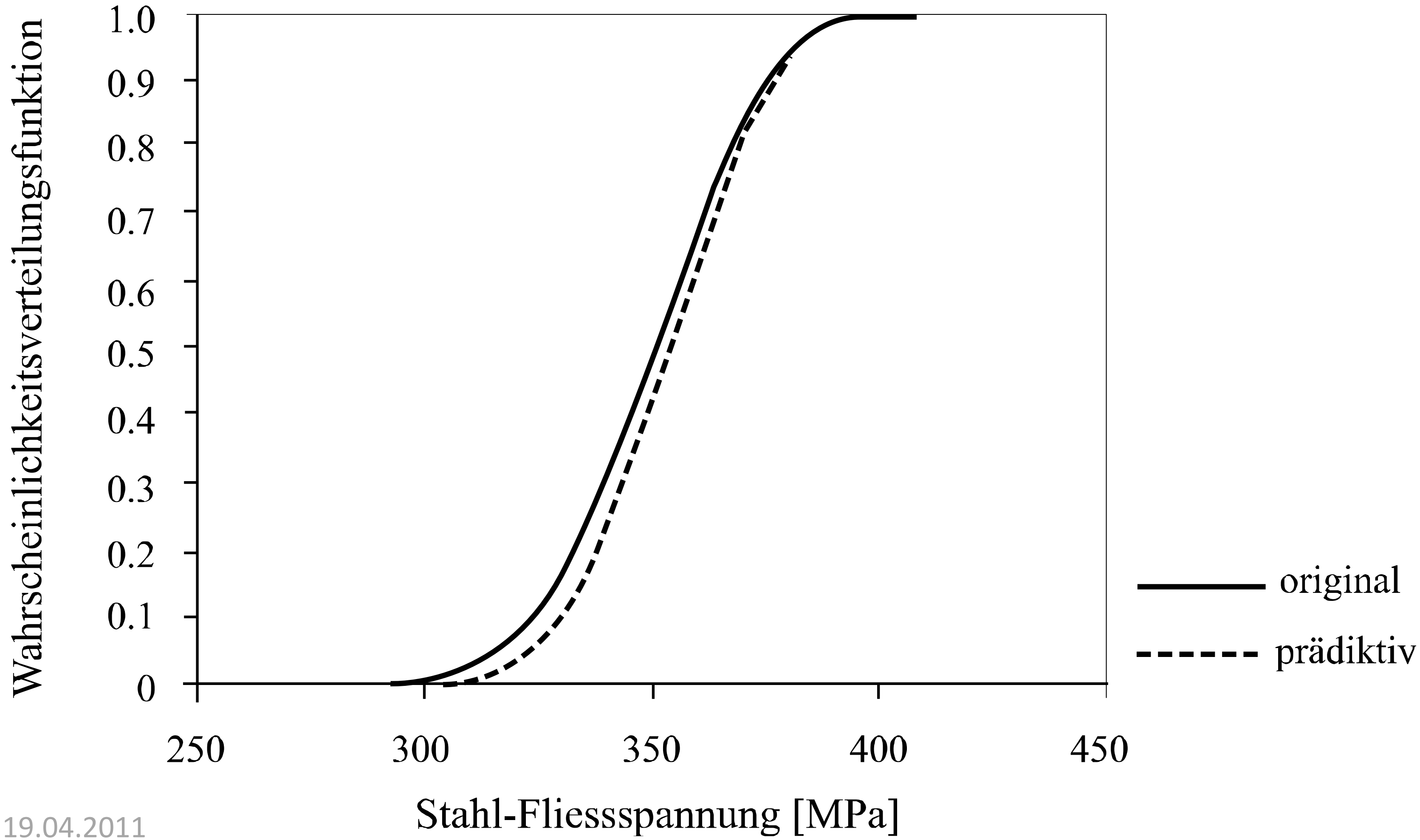
$$f_{f_y}\left(f_y \middle| \hat{\mathbf{f}}_y\right) = \frac{1}{\sqrt{2\pi}\sigma'''} \exp\left(-\frac{1}{2}\left(\frac{f_y - \mu''}{\sigma'''}\right)^2\right)$$

$$\sigma'''^2 = \sigma''^2 + \sigma_{f_y}^2$$

Die prädiktive Verteilung kombiniert unsere Informationen aus dem a priori Modell und den 5 Testergebnissen.

Bayes'sche Methoden der Parameterschätzung

Beispiel: Fliessspannung von Stahl



Bayes‘sche Regressionsanalyse

Im Ingenieurwesen werden oft verschiedene Arten von Modellen miteinander kombiniert:

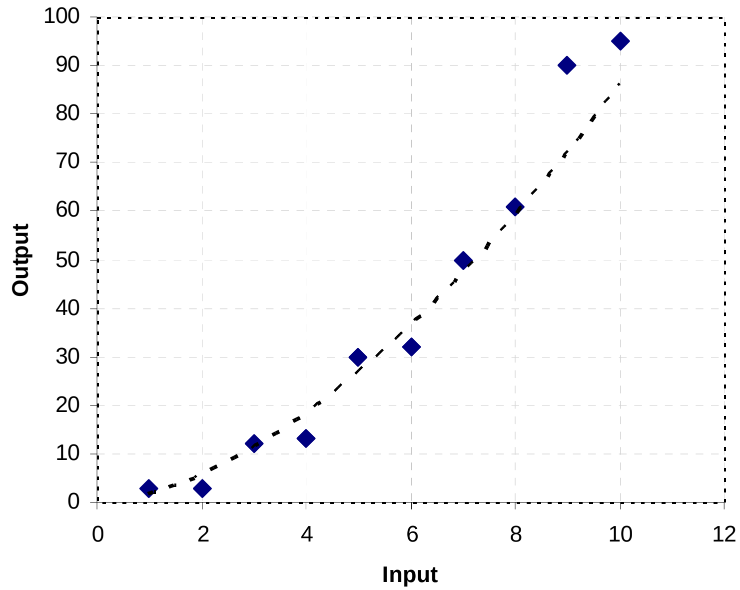
1) Physikalische Modelle

$$\frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(D(x,t)\frac{\partial C(x,t)}{\partial x}\right)$$

$$C(0,t) = C_s$$ Diffusionsgleichung

2) Empirische Modelle

$$y = f(x) = x^2$$ $\longrightarrow$



Bayes'sche Regressionsanalyse

Lineare Regression

Wir nehmen an, dass ein Zufallsvariable Y (Output) als lineare Funktion einer Zufallsvariablen X (Input) gegeben ist.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Aus n paarweisen Beobachtungen $(\hat{\mathbf{x}}, \hat{\mathbf{y}})^T$ möchten wir die Regressionsgerade schätzen.

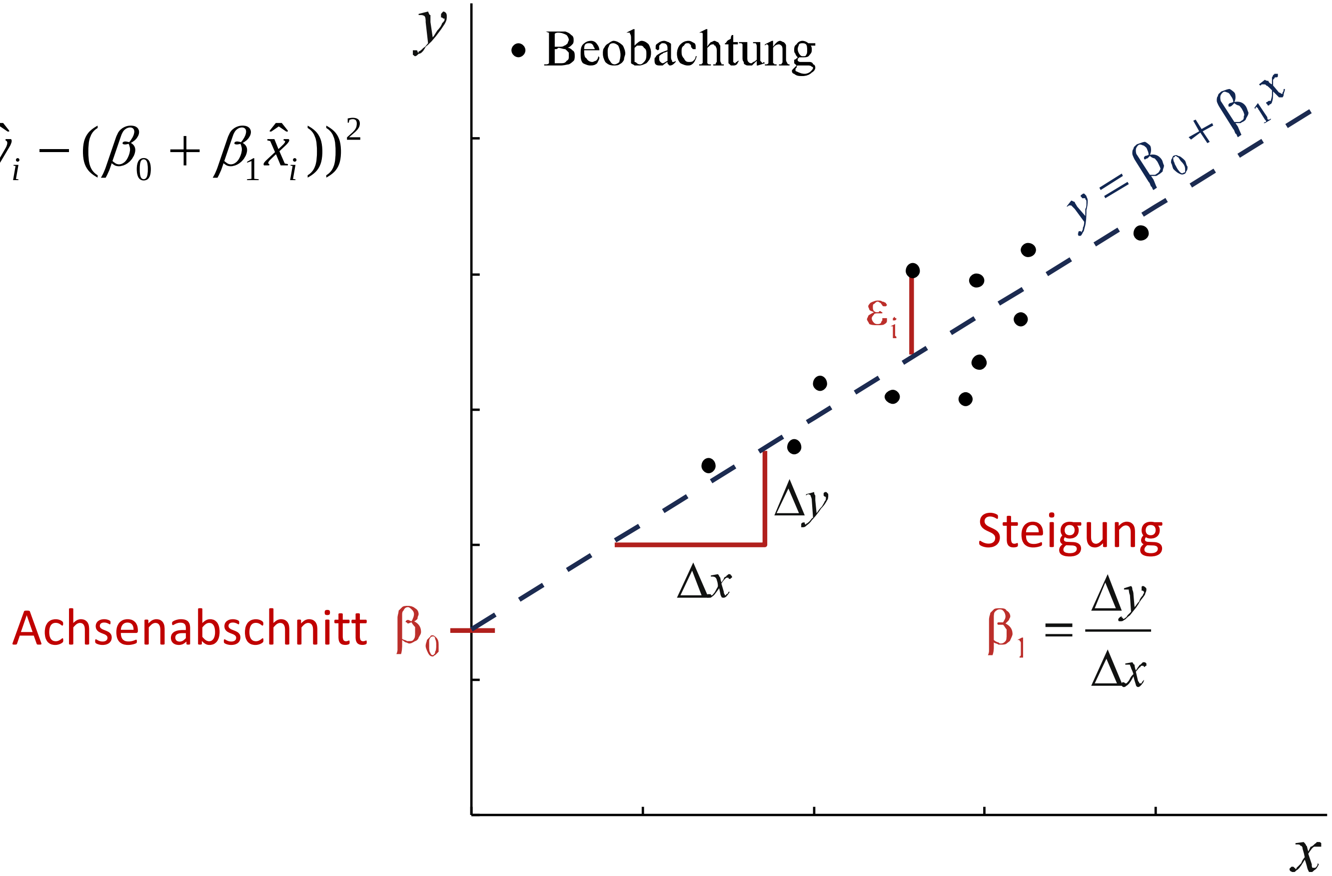
Wir nehmen an, dass der Residualwert (Fehler) ε einer Normalverteilung folgt: $\varepsilon \sim N(0, \sigma_\varepsilon)$

Oft kann eine nichtlineare Funktion durch Transformation linearisiert werden.

Bayes'sche Regressionsanalyse

Die Modellparameter β_0 und β_1 können mit der Methode der kleinsten Quadrate geschätzt werden.

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} (\hat{y}_i - (\beta_0 + \beta_1 \hat{x}_i))^2$$



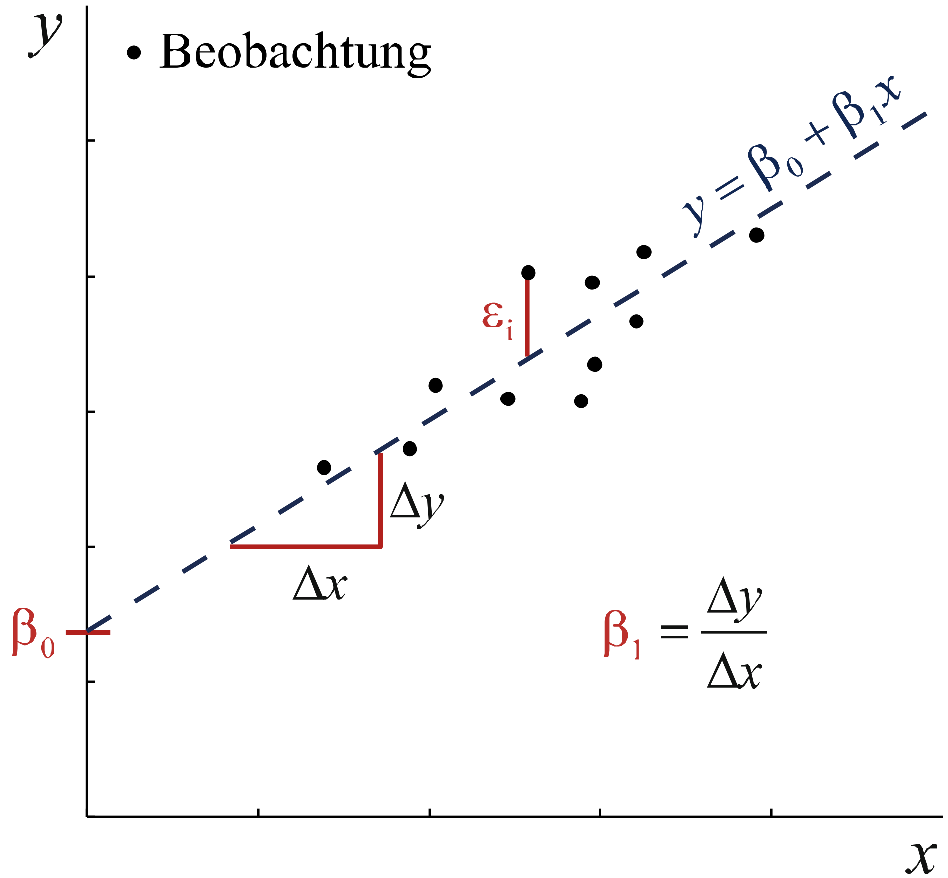
Bayes'sche Regressionsanalyse

Die Modellparameter β_0 und β_1 können mit der Methode der kleinsten Quadrate geschätzt werden.

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} (\hat{y}_i - (\beta_0 + \beta_1 \hat{x}_i))^2$$

Es ergibt sich folgendes Gleichungssystem:

$$\rightarrow \left| \begin{array}{l} \dfrac{\partial}{\partial \beta_0} \left(\sum_{i=1}^{n} (\hat{y}_i - (\beta_0 + \beta_1 \hat{x}_i))^2 \right) = 0 \\ \dfrac{\partial}{\partial \beta_1} \left(\sum_{i=1}^{n} (\hat{y}_i - (\beta_0 + \beta_1 \hat{x}_i))^2 \right) = 0 \end{array} \right|$$



Bayes‘sche Regressionsanalyse

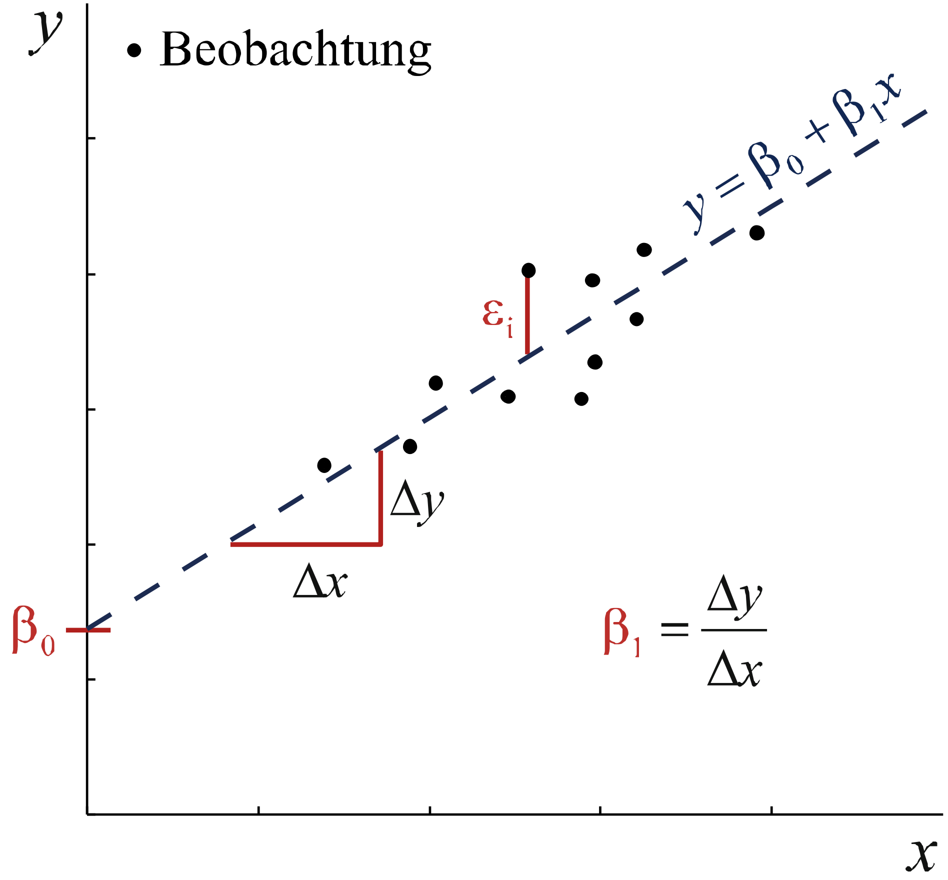
Die Modellparameter β_0 und β_1 können mit der Methode der kleinsten Quadrate geschätzt werden.

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} (\hat{y}_i - (\beta_0 + \beta_1 \hat{x}_i))^2$$

Lösung des Gleichungssystems:

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i - \beta_1 \cdot \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = \overline{y} - \overline{x}\beta_1$$

$$\beta_1 = \frac{\sum_{i=1}^{n} \hat{x}_i \hat{y}_i - \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} \hat{x}_i^2 - n \cdot (\overline{x})^2} = \frac{S_{XY}}{S_X}$$



Bayes'sche Regressionsanalyse

Die Unsicherheit des Regressionsmodells kann mit einer Normalverteilung für den Modelloutput Y abgebildet werden:

$$Y \sim N(\underbrace{\beta_0 + \beta_1 X}_{\text{Mittelwert}};\underbrace{\sigma_\varepsilon}_{\text{Standardabweichung}})$$

Schätzung der Modellparameter:

$$\beta_0 = \overline{y} - \overline{x}\beta_1 \qquad \beta_1 = \frac{S_{XY}}{S_X}$$

Standardabweichung des Fehlers:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-k}}$$

n Anzahl Beobachtungen
k Anzahl Parameter in $\boldsymbol{\beta}$

Die gleichen Ergebnisse erhält man auch mit der Maximum Likelihood Methode

Bayes'sche Regressionsanalyse

Das Gleichungssystem für die Schätzung der Modellparameter lässt sich auch in Matrix-Schreibweise formulieren:

$$\hat{\mathbf{X}}^T \hat{\mathbf{y}} = \hat{\mathbf{X}}^T \hat{\mathbf{X}} \boldsymbol{\beta} \longrightarrow \boldsymbol{\beta} = (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{y}}$$

Hierin ist:

$$\boldsymbol{\beta} = (\beta_0, \beta_1)^T$$

$$\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, .., \hat{y}_n)^T$$

$$\hat{\mathbf{X}} = \begin{pmatrix} 1 & \hat{x}_1 \\ 1 & \hat{x}_2 \\ \vdots & \vdots \\ 1 & \hat{x}_n \end{pmatrix}$$

Diese Schreibweise ist insbesondere nützlich, um die Regressionsanalyse zu programmieren.

Bayes‘sche Regressionsanalyse

Die Matrixschreibweise erleichtert auch die Verallgemeinerung auf den mehrdimensionalen Fall:

$$y = \beta_0 + \sum_{j=1}^{r} x_j \beta_j + \varepsilon \quad \longrightarrow \quad \boldsymbol{\beta} = (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{y}}$$

Dann ist:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)^T$$

$$\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^T$$

$$\hat{\mathbf{X}} = \begin{pmatrix} 1 & \hat{x}_{12} & \cdots & \hat{x}_{1r} \\ 1 & \hat{x}_{21} & \cdots & \hat{x}_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \hat{x}_{n1} & \cdots & \hat{x}_{nr} \end{pmatrix}$$

Bayes'sche Regressionsanalyse

Die statistische Unsicherheit des Regressionmodells lässt sich bei Verwendung der Matrixschreibweise wie folgt quantifizieren:

$$\sigma_\varepsilon^2 = \frac{1}{n-k}\left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right)^T \left(\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}\right)$$

mit n Anzahl Beobachtungen
k Anzahl Parameter in $\boldsymbol{\beta}$

Varianz des Fehlers

Daraus ergibt sich die Unsicherheit der Parameterschätzung:

$$Cov\left[\boldsymbol{\beta}\right] = \sigma_\varepsilon^2 \mathbf{V}_\beta \quad \text{mit} \quad \mathbf{V}_\beta = \left(\hat{\mathbf{X}}^T \hat{\mathbf{X}}\right)^{-1}$$

Kovarianz-Matrix der Parameter

Gegeben $\mathbf{X}$ und $\boldsymbol{\beta}$ folgt Y einer Normalverteilung:

$$E\left[Y \middle| \boldsymbol{\beta}; \mathbf{X}\right] = \boldsymbol{\beta}\mathbf{X} \qquad Var\left[Y \middle| \boldsymbol{\beta}; \mathbf{X}\right] = \sigma_\varepsilon^2$$

Bayes'sche Regressionsanalyse

Aktualisieren der Regressionskoeffizienten

Ein bestehendes Regressionsmodell kann mit neuen Daten aktualisiert werden. Hierfür benötigen wir

- Informationen über das *a priori* Modell:

$$\boldsymbol{\beta}' = (\beta_0', \beta_1')^T \qquad \mathbf{V}_\beta' = \frac{Cov[\boldsymbol{\beta}']}{(\sigma_\varepsilon')^2}$$

- Informationen zu den neuen Beobachtungen:

$$\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, .., \hat{y}_n)^T \qquad \hat{\mathbf{X}} = \begin{pmatrix} 1 & \hat{x}_1 \\ 1 & \hat{x}_2 \\ \vdots & \vdots \\ 1 & \hat{x}_n \end{pmatrix}$$

Bayes'sche Regressionsanalyse

Die Parameter des *a posteriori* Modells ergeben sich wie folgt:

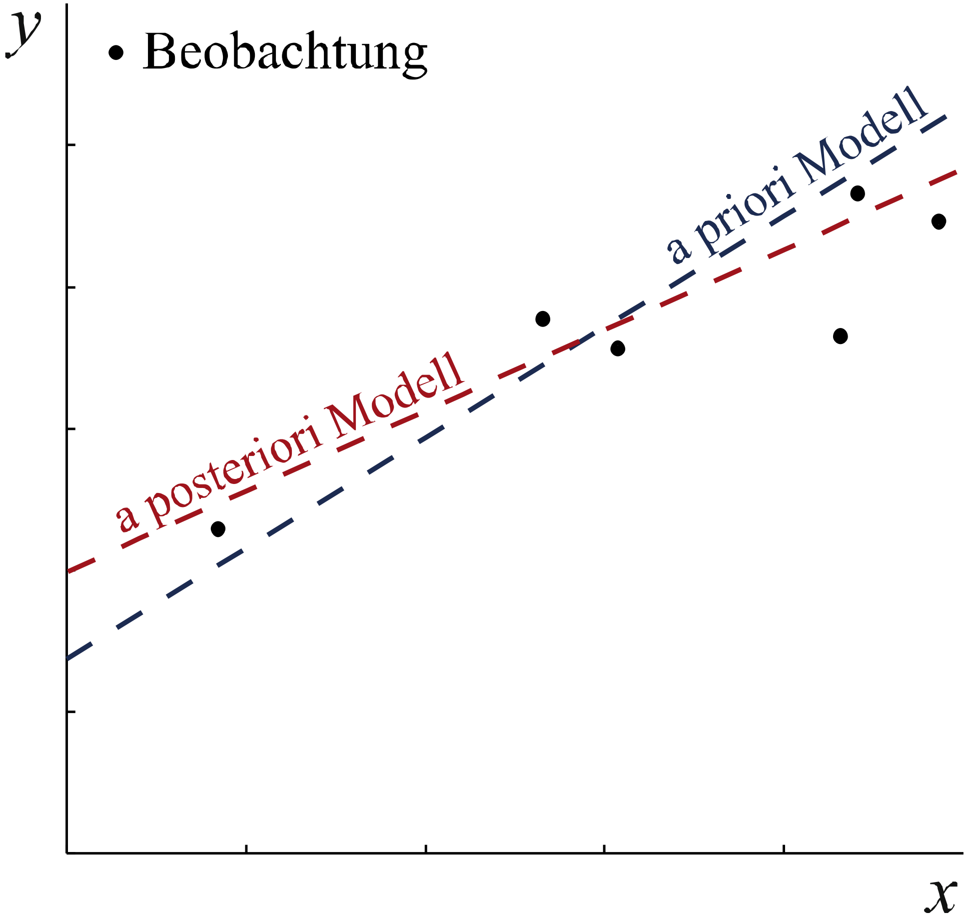
A posteriori Parameter

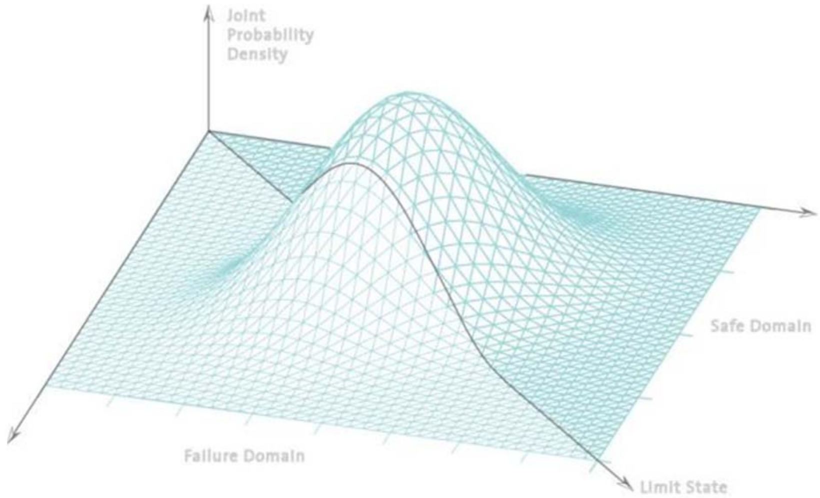
$$\boldsymbol{\beta}'' = \mathbf{V}''_{\beta}\left(\left(\mathbf{V}'_{\beta}\right)^{-1}\boldsymbol{\beta}' + \hat{\mathbf{X}}^{T}\hat{\mathbf{y}}\right)$$

A priori Parameter

Daten

Hierin ist $(\mathbf{V}''_{\beta})^{-1} = (\mathbf{V}'_{\beta})^{-1} + \hat{\mathbf{X}}^{T}\hat{\mathbf{X}}$





Statistik und Wahrscheinlichkeitsrechnung

Jochen Köhler