

# Statistik und Wahrscheinlichkeitsrechnung

## 10. Vorlesung

Prof. Dr. Michael Havbro Faber

## Dienstag, 4. Mai 2010: Übung

Raumaufteilung für die Übung am Dienstag:

Übung von Mathias und Markus: HIL E 1

Übung von Gerhard und Katharina: HIL E 6

Start: 8 Uhr

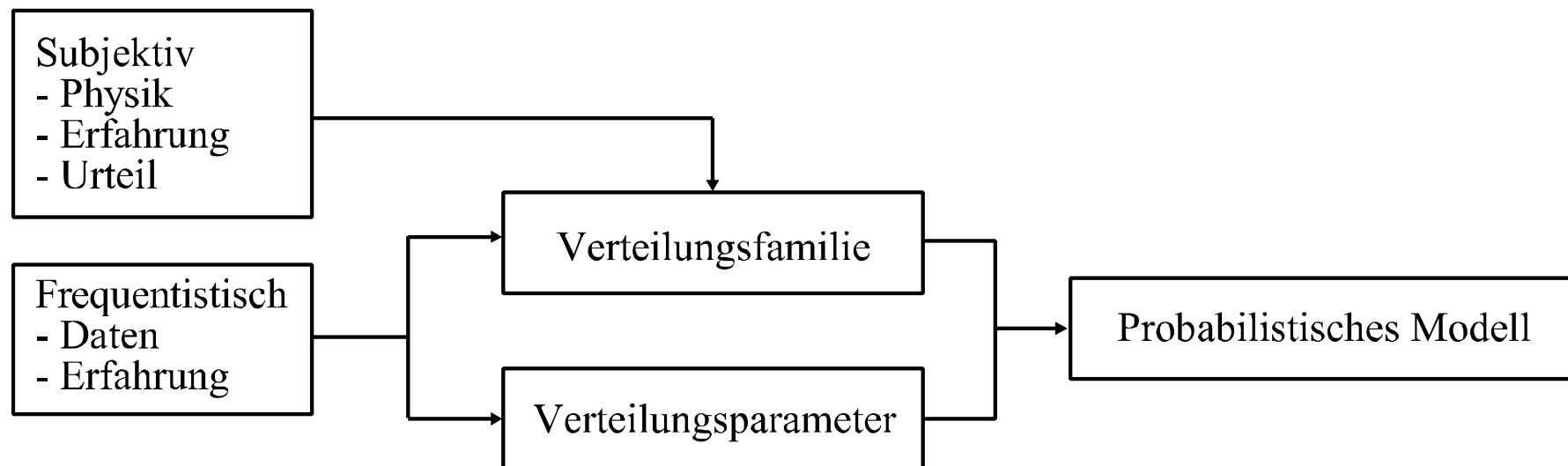
# Inhalte der heutigen Vorlesung

- 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



# Modellierung von Unsicherheiten - Übersicht

Knowledge is experience

Data is just information

A. Einstein

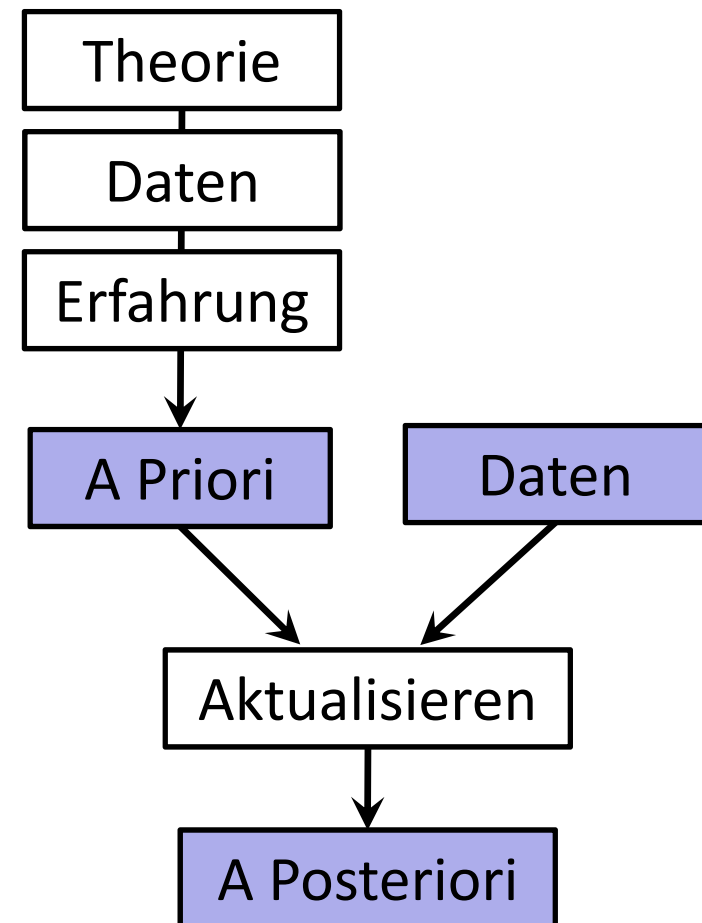
# Bayes'sche Methoden der Parameterschätzung

Bayes'sche Methoden der Parameterschätzung ermöglichen die gleichzeitige Berücksichtigung von

*a priori* (subjektiver und/oder frequentistischer) Information

und

zusätzlicher frequentistischer Information (Versuchsergebnisse oder Beobachtungen)



# Bayes'sche Methoden der Parameterschätzung

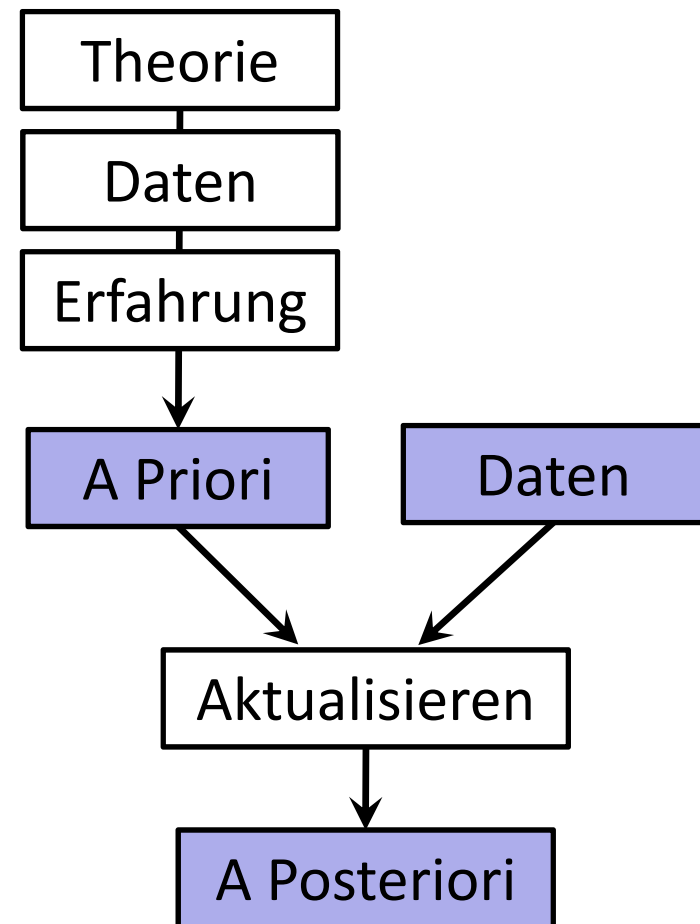
Bayes'sche Methoden der Parameterschätzung ermöglichen die gleichzeitige Berücksichtigung von

*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:

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



# Bayes'sche Methoden der Parameterschätzung

Die Likelihood der Beobachtungen berechnet sich zu:

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

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

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

Nun können wir die Dichtefunktion für  $\boldsymbol{\theta}$  aktualisieren. So erhalten wir die *prädiktive Dichtefunktion*:

$$f_X(x|\mathbf{x}) = \int f_X(x|\boldsymbol{\theta})f_{\boldsymbol{\theta}}''(\boldsymbol{\theta}|\hat{\mathbf{x}})d\boldsymbol{\theta}$$

# 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

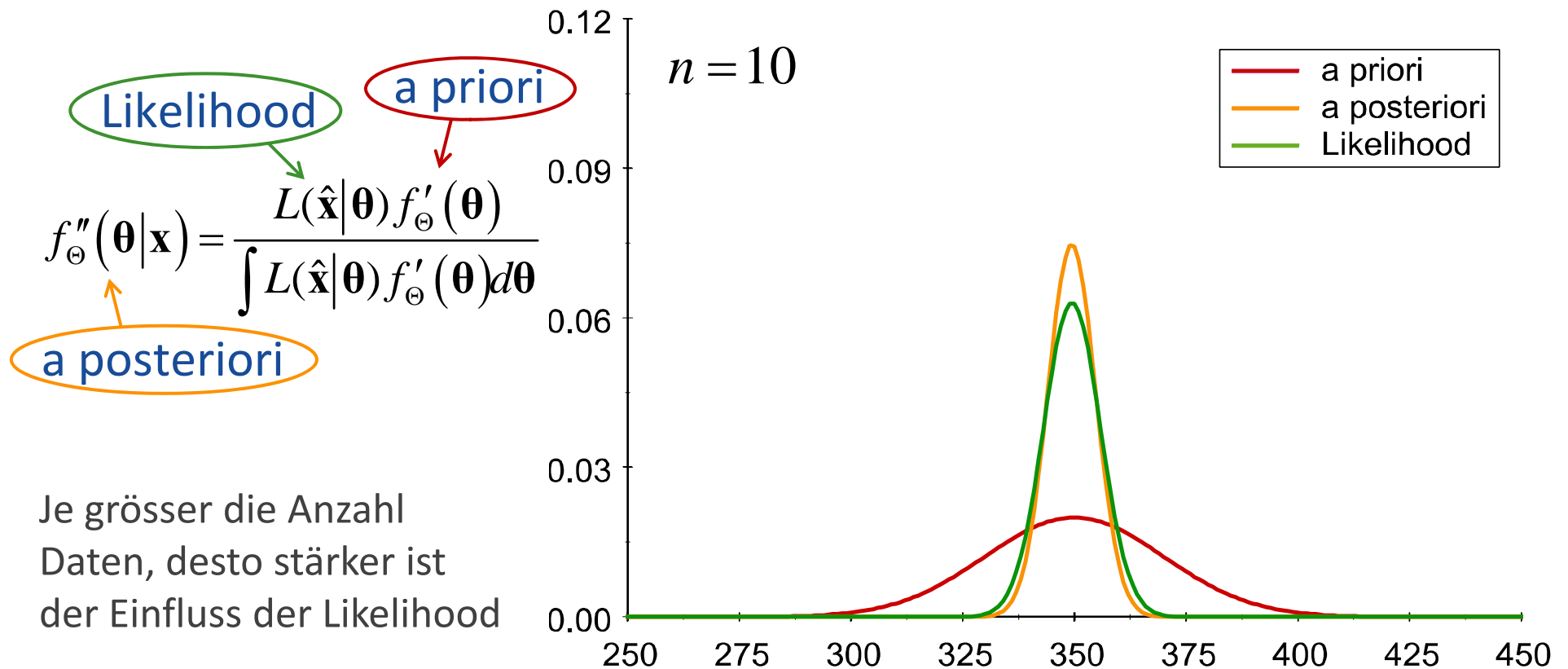
$$f_{\Theta}''(\boldsymbol{\theta}|\mathbf{x}) = \frac{L(\hat{\mathbf{x}}|\boldsymbol{\theta})f'_{\Theta}(\boldsymbol{\theta})}{\int L(\hat{\mathbf{x}}|\boldsymbol{\theta})f'_{\Theta}(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

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

- Mittelwert der Fließgrenze von Stahl
  - Standardabweichung einer Schneelast
- statistische Unsicherheit

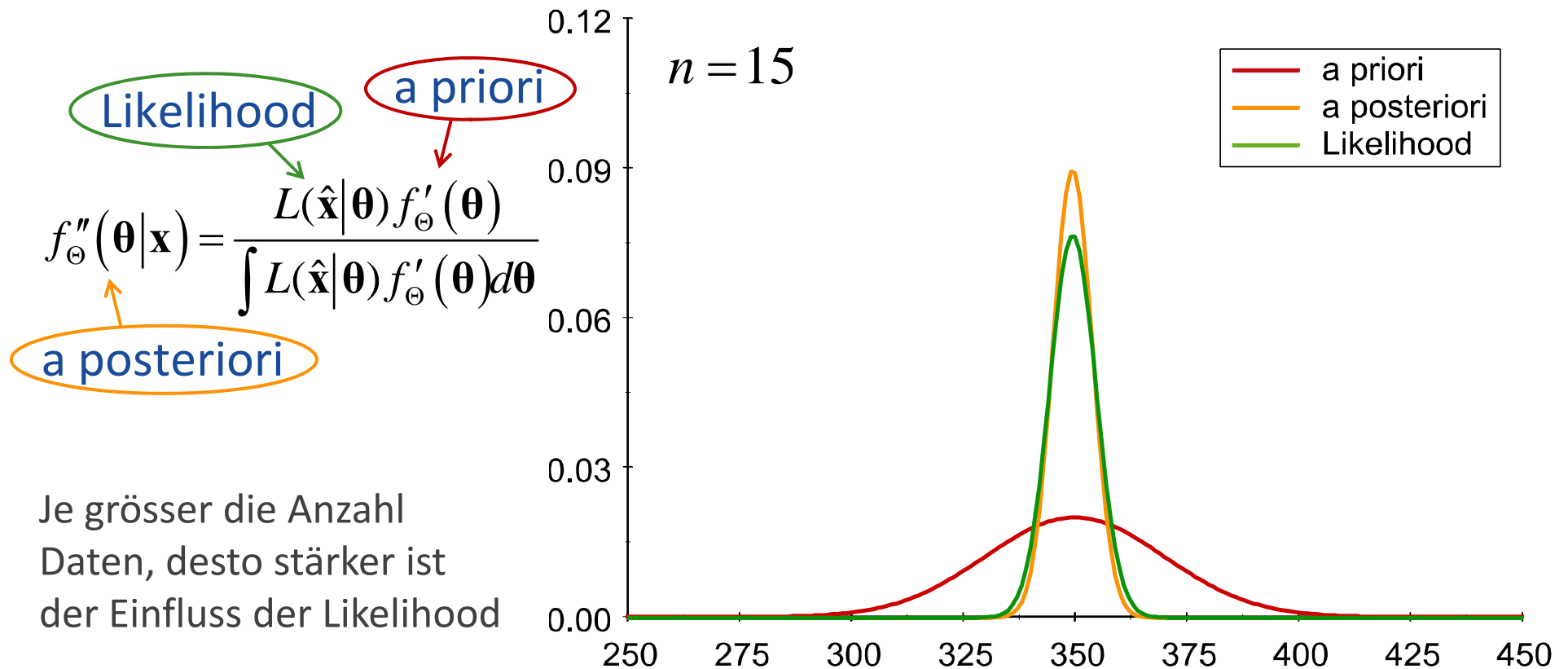
# Bayes'sche Methoden der Parameterschätzung

Illustration: Aktualisieren einer Dichtefunktion



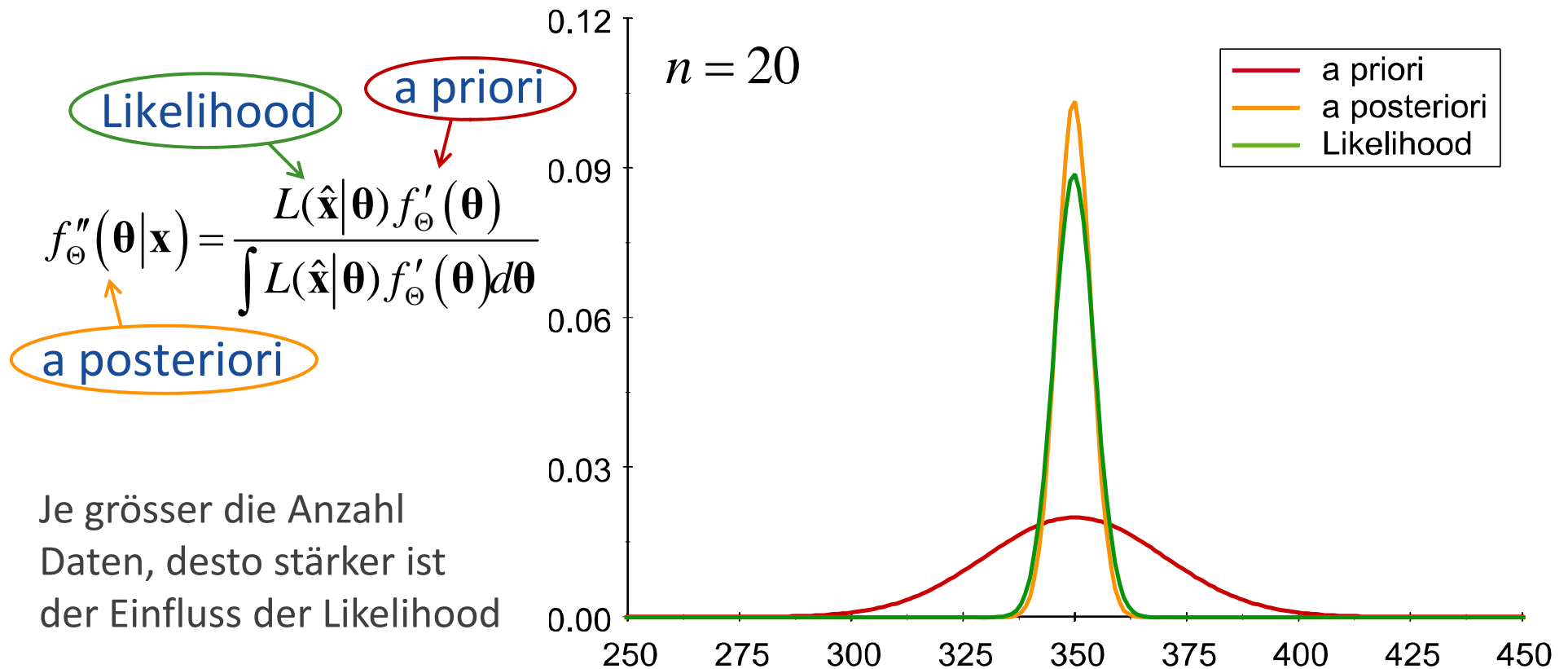
# Bayes'sche Methoden der Parameterschätzung

Illustration: Aktualisieren einer Dichtefunktion



# Bayes'sche Methoden der Parameterschätzung

Illustration: Aktualisieren einer Dichtefunktion



# Bayes'sche Methoden der Parameterschätzung

## Beispiel: Fließspannung von Stahl

Wir nehmen an, die Fließspannung  $f_y$  eines Baustahls in  $MPa$  sei normalverteilt mit Erwartungswert  $\mu_{f_y}$  und Standardabweichung  $\sigma_{f_y}$ :

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

Die Standardabweichung sei bekannt, der Erwartungswert unsicher:

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

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 Standardabweichung und unbekanntem, normalverteiltem Erwartungswert ist die *a posteriori* Verteilung des Erwartungswerts ebenfalls normalverteilt:

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



# Bayes'sche Methoden der Parameterschätzung

## Beispiel: Fließspannung von Stahl

Bei einer normalverteilten Zufallsvariable mit bekannter Standardabweichung 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:

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

# Bayes'sche Methoden der Parameterschätzung

## Beispiel: Flie遢spannung von Stahl

Bei einer normalverteilten Zufallsvariable mit bekannter Standardabweichung 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:

$$n' = 3.06 \quad \mu'' = 353.22 \text{ MPa} \quad \sigma'' = 6.16 \text{ MPa}$$

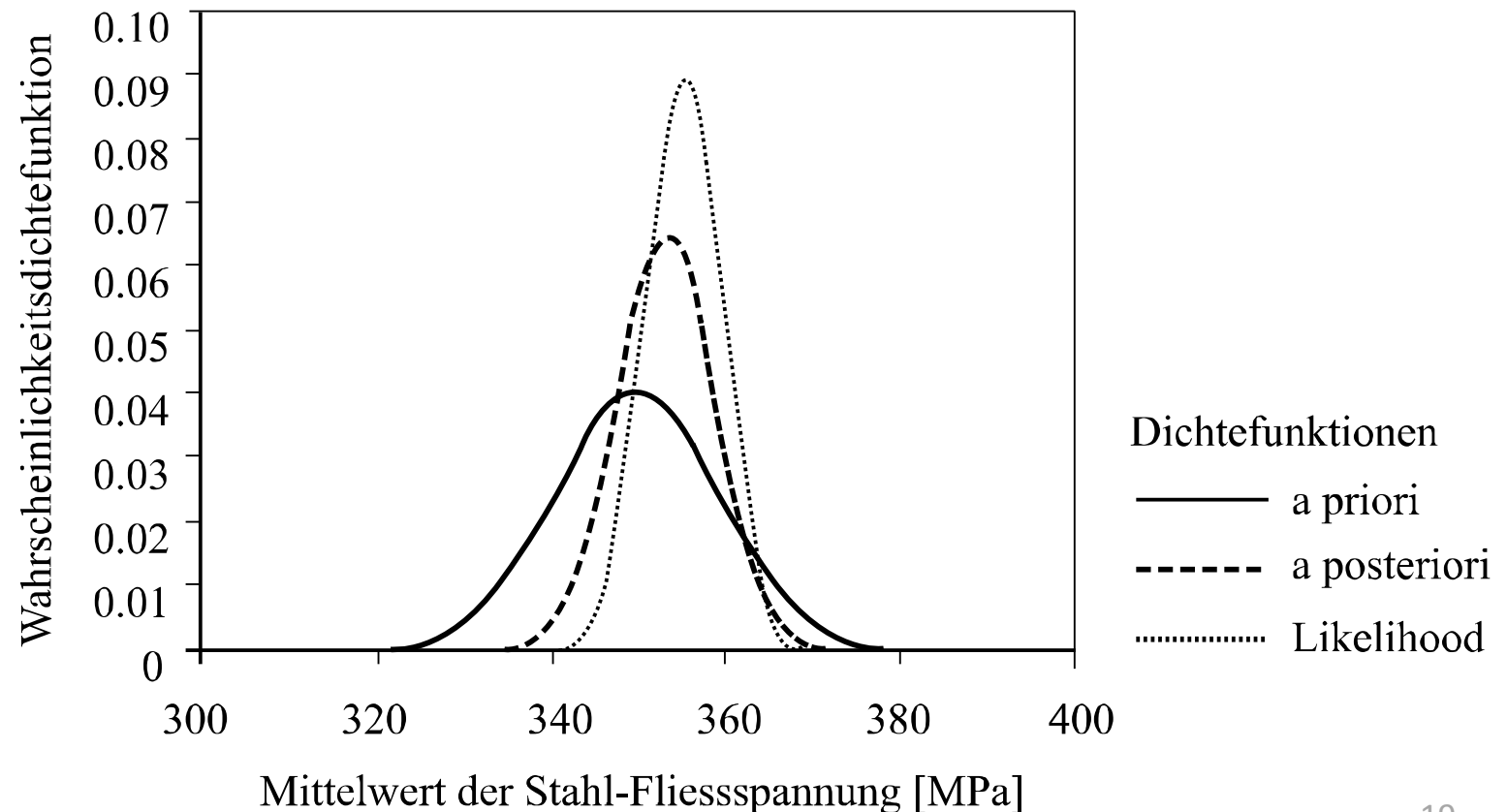
# Bayes'sche Methoden der Parameterschätzung

## 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 *prädiktive* Wahrscheinlichkeitsdichtefunktion für die Stahl-Fließspannung bestimmen:

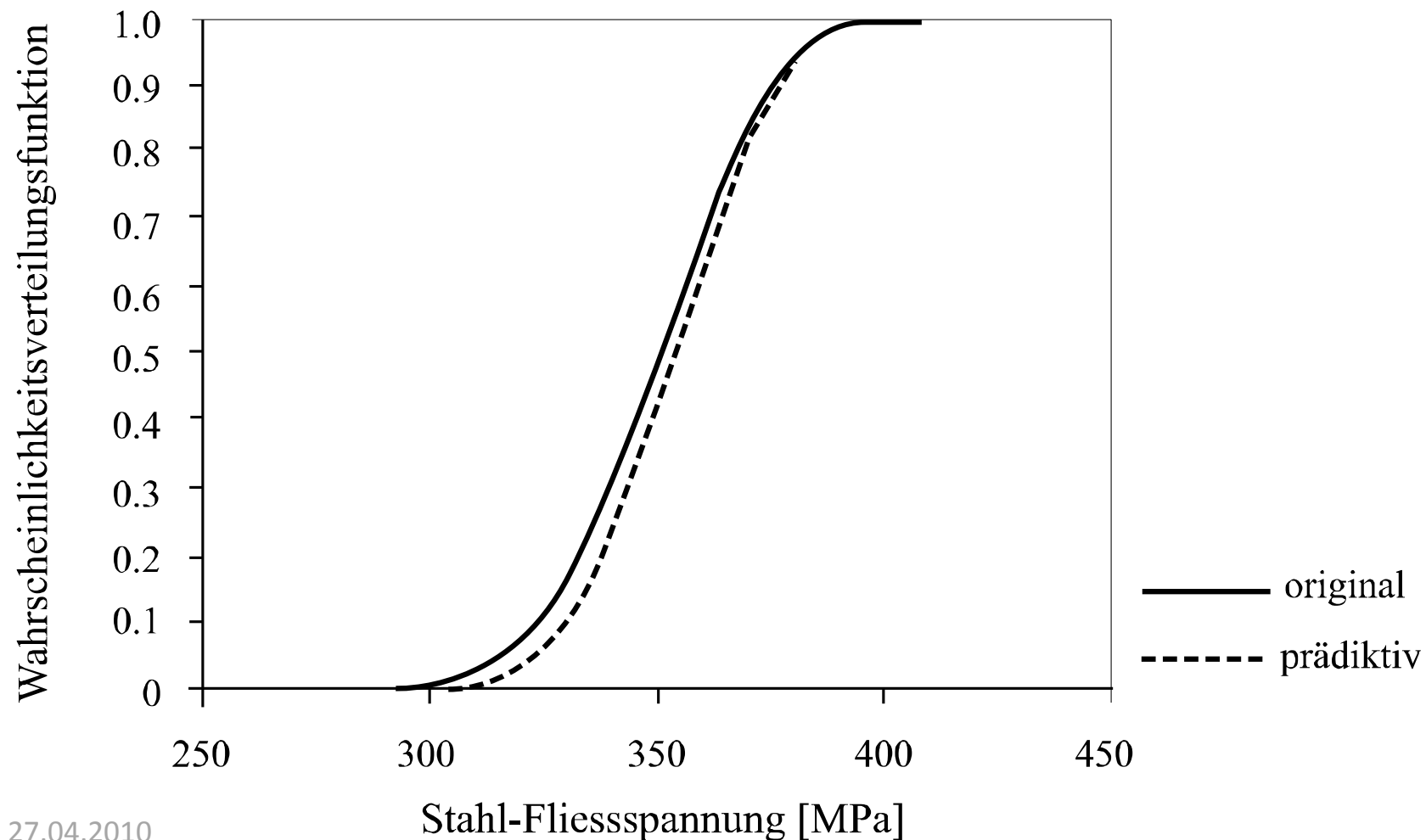
$$f_{f_y} \left( f_y \mid \hat{\mathbf{f}}_y \right) = \frac{1}{\sqrt{2\pi}\sigma'''} \exp \left( -\frac{1}{2} \left( \frac{f_y - \mu''}{\sigma'''} \right)^2 \right)$$

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

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:

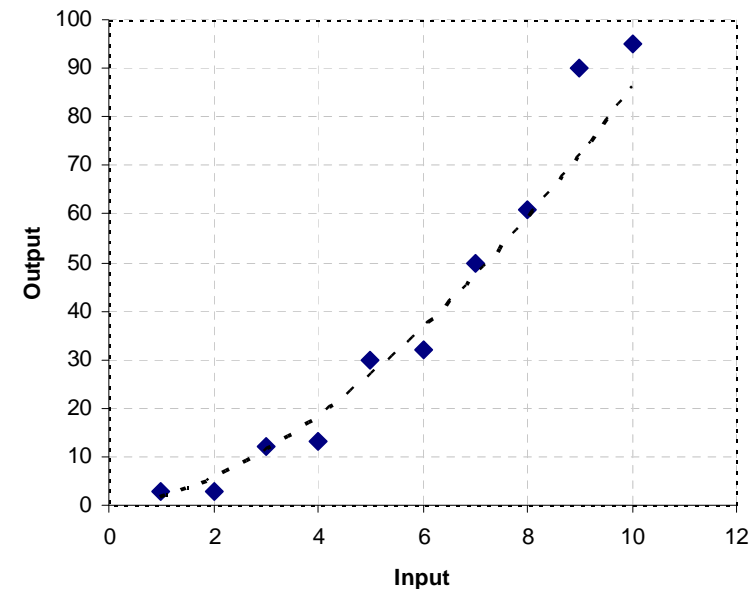
## 1) Physikalische Modelle

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

$$C(0,t) = C_s \quad \text{Diffusionsgleichung}$$

## 2) Empirische Modelle

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



# Bayes'sche Regressionsanalyse

## Lineare Regression

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

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

$$y = \beta_0 + \beta_1 x + \varepsilon$$

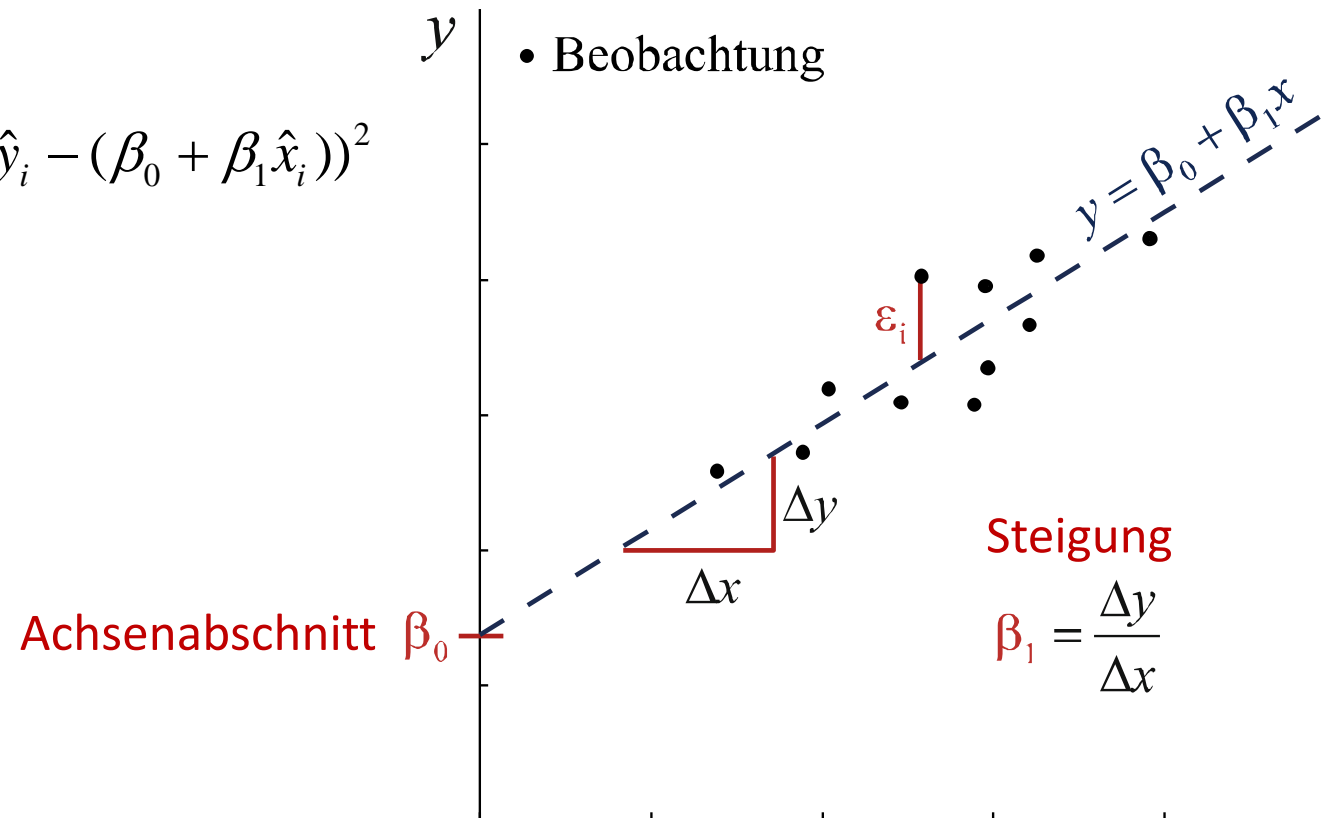
Weiter nehmen wir an, dass der Residualwert (Fehler)  $\varepsilon$  einer Normalverteilung folgt:  $\varepsilon \sim N(0, \sigma_\varepsilon)$

Oft kann eine nichtlineare Funktion durch Transformation linearisiert werden.

# Bayes'sche Regressionsanalyse

Die Modellparameter  $\beta_0$  und  $\beta_1$  können mit der Methode der kleinsten Quadrate geschätzt werden.

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





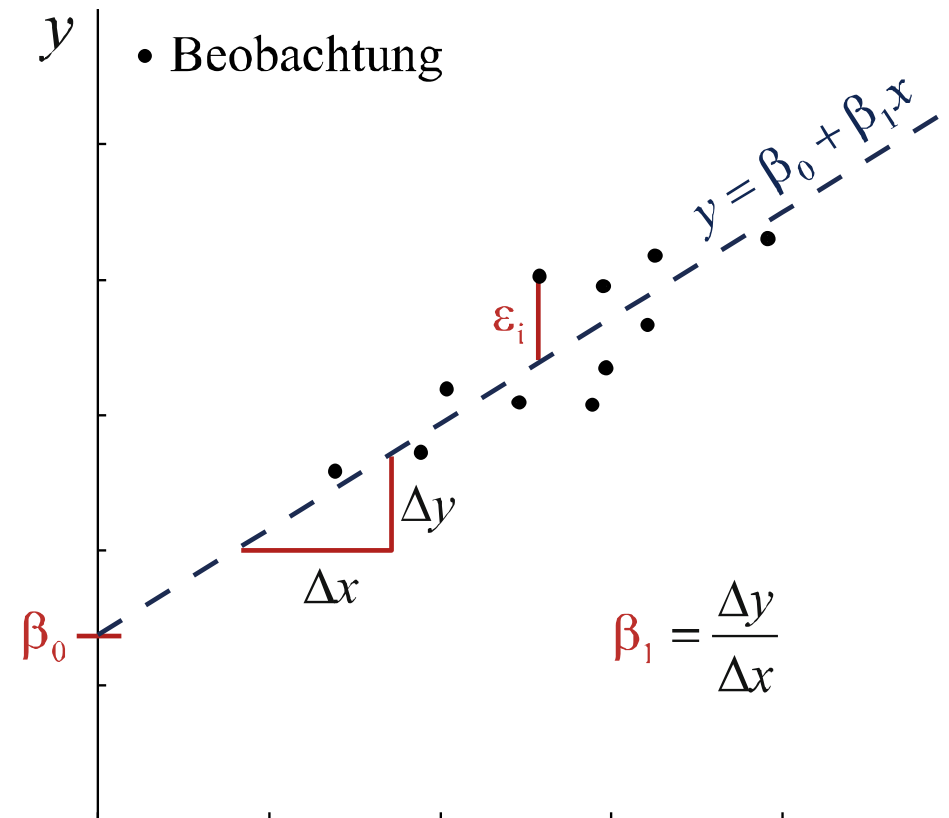
# Bayes'sche Regressionsanalyse

Die Modellparameter  $\beta_0$  und  $\beta_1$  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:

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



# Bayes'sche Regressionsanalyse

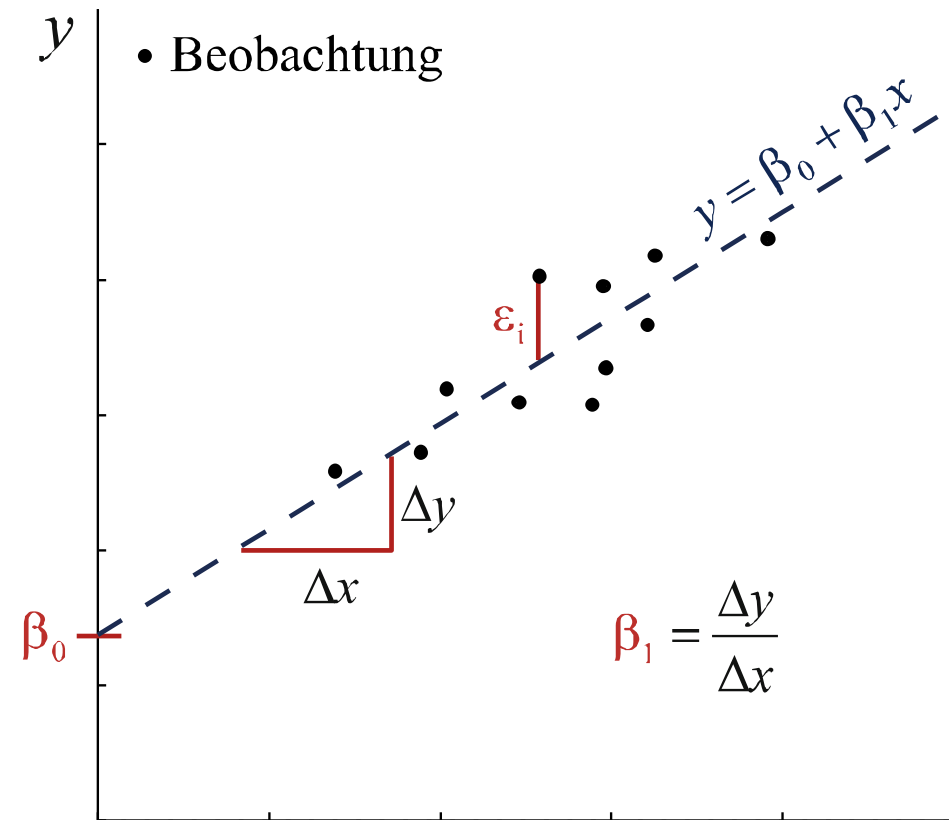
Die Modellparameter  $\beta_0$  und  $\beta_1$  können mit der Methode der kleinsten Quadrate geschätzt werden.

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

Lösung des Gleichungssystems:

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

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



# 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:

$$\beta_0 = \bar{y} - \bar{x}\beta_1 \quad \beta_1 = \frac{S_{XY}}{S_X}$$

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:

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

Hierin ist:  $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$   
 $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$

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

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

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

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

# Bayes'sche Regressionsanalyse

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

$$\sigma_{\varepsilon}^2 = \frac{1}{n-k} (\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta})^T (\hat{\mathbf{y}} - \hat{\mathbf{X}}\boldsymbol{\beta}) \quad \begin{array}{l} \text{mit } n \text{ Anzahl Beobachtungen} \\ \quad k \text{ Anzahl Parameter in } \boldsymbol{\beta} \end{array}$$

↖  
Varianz des Fehlers

Daraus ergibt sich die Unsicherheit der Parameterschätzung:

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

↖  
Kovarianz-Matrix der Parameter

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

$$E[Y | \boldsymbol{\beta}; \mathbf{X}] = \boldsymbol{\beta}\mathbf{X} \quad \text{Var}[Y | \boldsymbol{\beta}; \mathbf{X}] = \sigma_{\varepsilon}^2$$

# Bayes'sche Regressionsanalyse

## Aktualisieren der Regressionskoeffizienten

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

- Informationen über das *a priori* Modell:

$$\boldsymbol{\beta}' = (\beta'_0, \beta'_1)^T \quad \mathbf{V}'_{\beta} = \frac{\text{Cov}[\boldsymbol{\beta}']}{(\sigma'_{\varepsilon})^2}$$

- Informationen zu den neuen Beobachtungen:

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

# Bayes'sche Regressionsanalyse

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

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

*A posteriori* Parameter  
*A priori* Parameter  
 Daten

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

## ACHTUNG:

Im Skript sind diese Gleichungen falsch angegeben –  
Korrekturseiten sind auf der Homepage

(betrifft Gleichungen E.45 und E.46, sowie das Beispiel E.4)



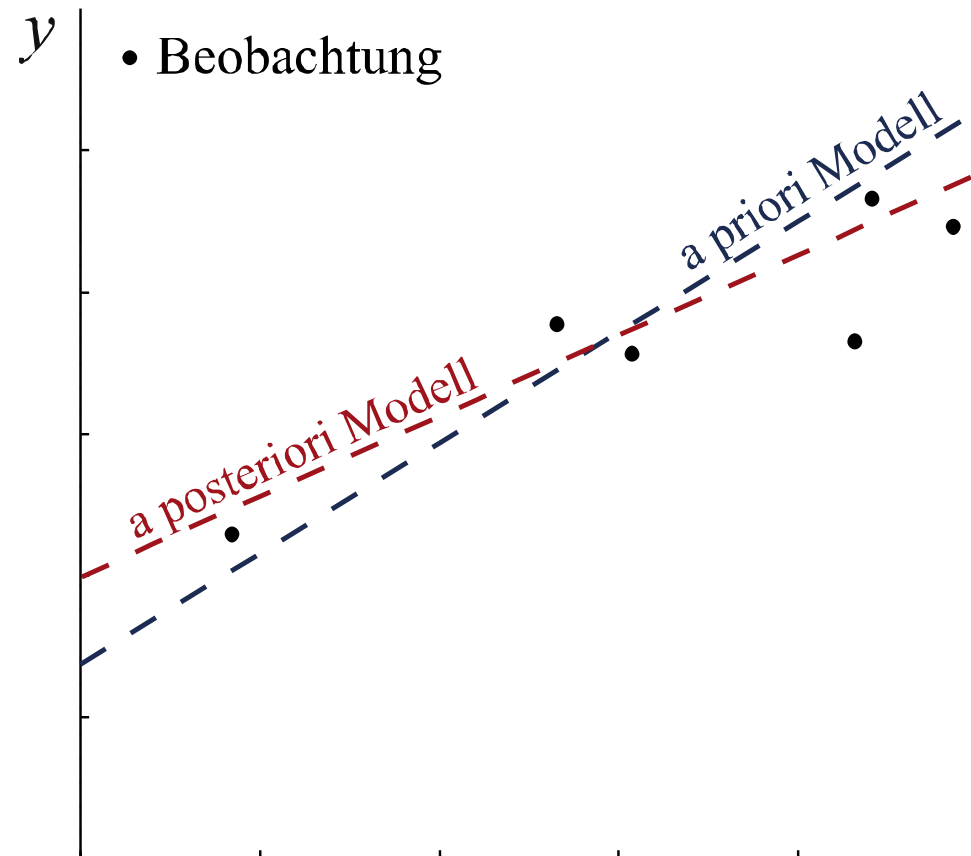
# Bayes'sche Regressionsanalyse

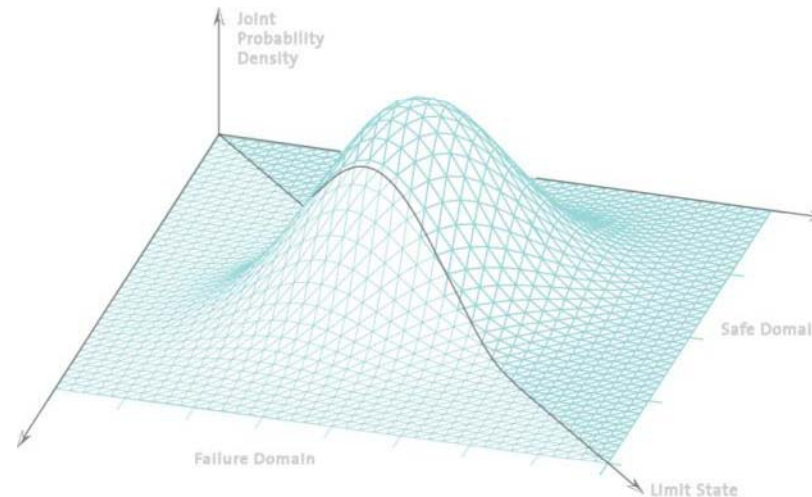
Die Parameter des *a posteriori* Modells ergeben sich wie folgt:

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

↙ *A posteriori* Parameter  
↖ *A priori* Parameter  
↗ Daten

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





# Statistik und Wahrscheinlichkeitsrechnung

Prof. Dr. Michael H. Faber