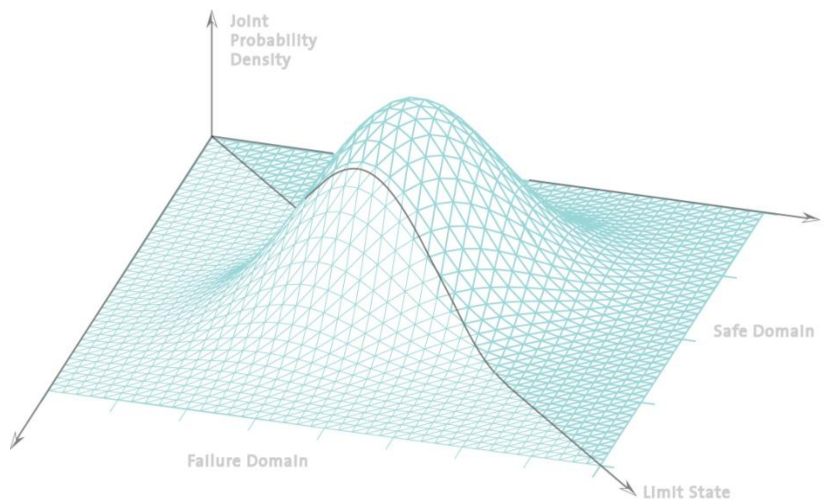


Statistik und Wahrscheinlichkeitsrechnung

Mathias Graf

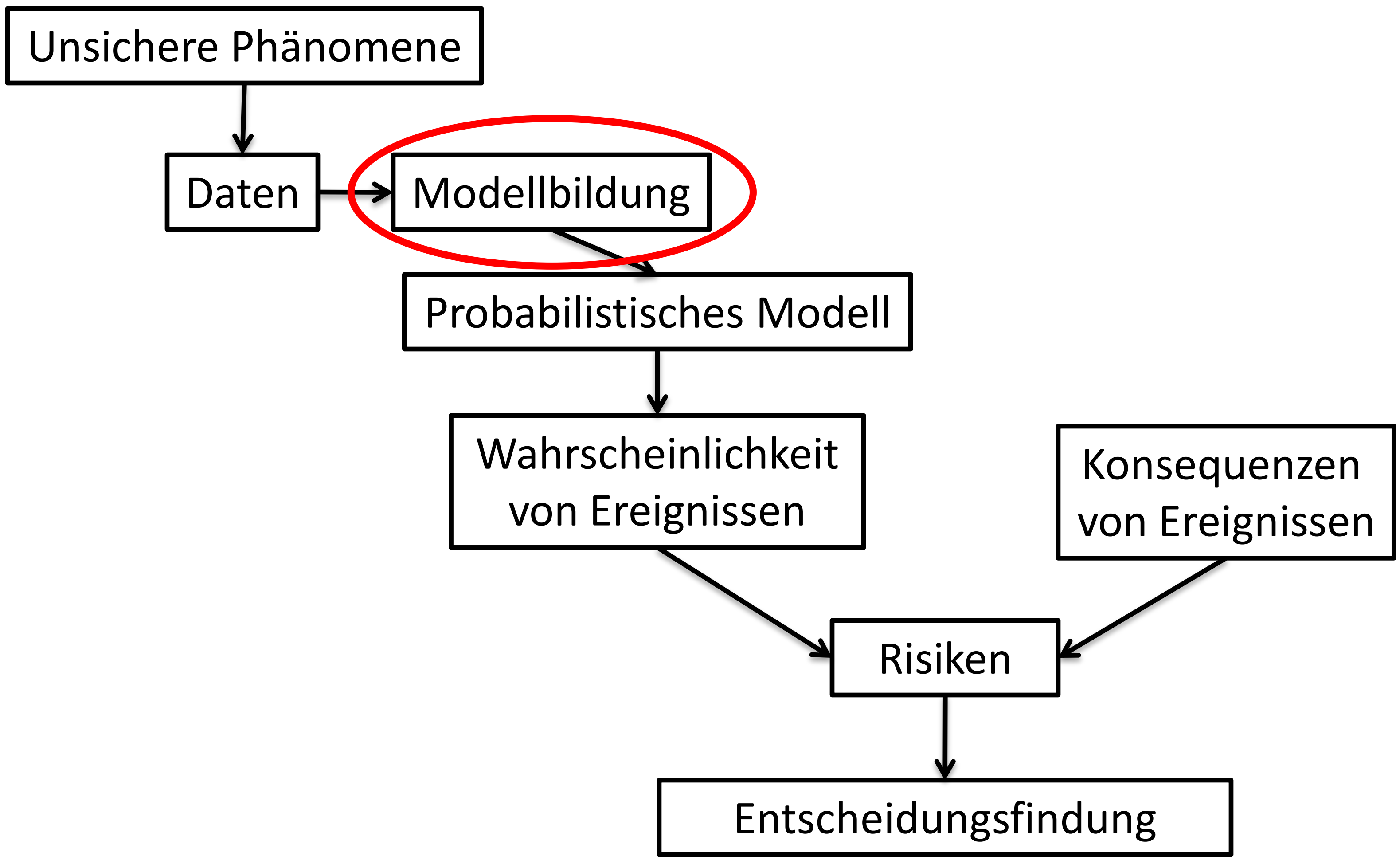
Inhalt der heutigen Vorlesung

- Auswahl einer Verteilungsfunktion: Wahrscheinlichkeitspapier

- Schätzung und Modellentwicklung:
  - Schätzung der Verteilungsparameter
  - Methode der Momente
  - Methode der Maximum Likelihood

Der Enscheidungskontext!

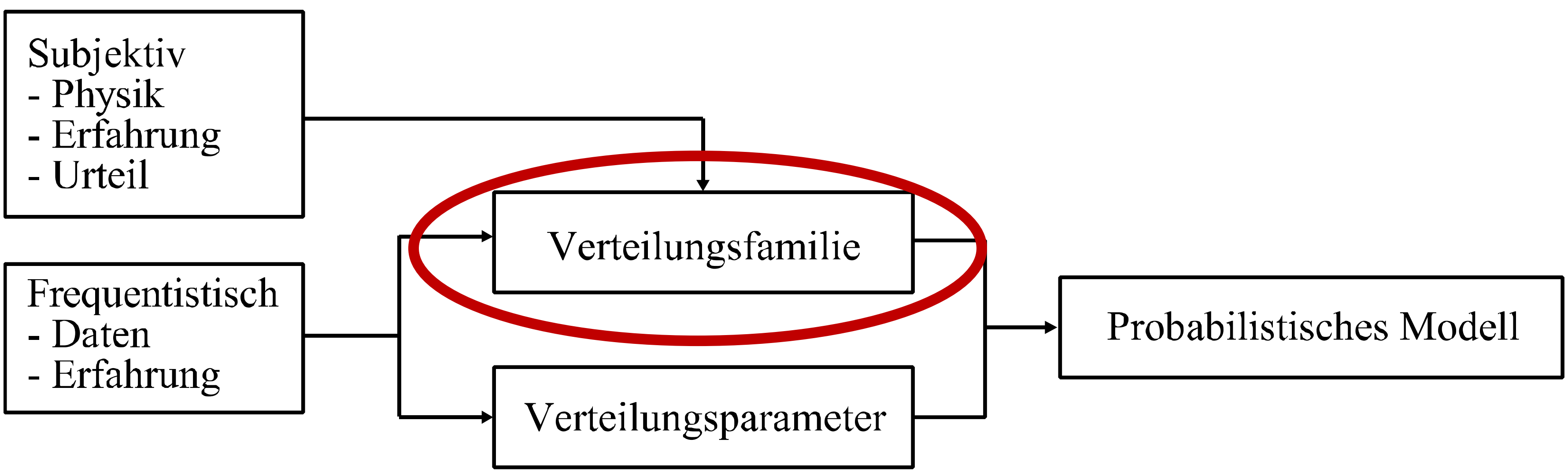
- Wieso Unsicherheitenmodellierung?



Schätzung und Modellentwicklung - Übersicht

Wenn man Modelle im Ingenieurbereich entwickeln möchte, müssen unterschiedliche Typen von Informationen herangezogen werden.

- subjektive Informationen
- frequentistische Informationen



Schätzung und Modellentwicklung

Auswahl von Wahrscheinlichkeitsverteilungsfunktionen

Grundsätzlich müssen Verteilungsfunktionen für Zufallsvariablen oder –prozesse auf Basis folgender Punkte ausgewählt werden:

Frequentistische Information:	Daten
Physikalische Argumente:	Verständnis Ingenieurproblemstellungen

Der klassische Ansatz ist folgender:

1. Bestimmen einer Hypothese für eine Wahrscheinlichkeitsverteilungsfamilie.
2. Schätzen der Funktionsparameter der bestimmten Wahrscheinlichkeitsverteilung.
3. Durchführen eines statistischen Tests, um die Hypothese abzulehnen oder zu akzeptieren.

Schätzung und Modellentwicklung

Auswahl von Wahrscheinlichkeitsverteilungsfunktionen

Im Ingenieurwesen tritt häufig der Fall ein, dass die verfügbaren Daten zu spärlich sind, um einen Hypothesentest für eine gegebene Wahrscheinlichkeitsverteilung durchzuführen – zumindest mit einer vernünftigen Signifikanz.

Deshalb ist ein einheitliches Vorgehen sehr wichtig:

Zuerst werden physikalische Argumente herangezogen, um eine passende Verteilung zu identifizieren.

Darauf aufbauend wird überprüft, ob die zur Verfügung stehenden Daten der gewählten Verteilungsfunktion widersprechen.

Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Ein Wahrscheinlichkeitspapier ist so skaliert, dass eine bestimmte Funktion beim Aufzeichnen auf dieses Papier die Form einer geraden Linie erhält.

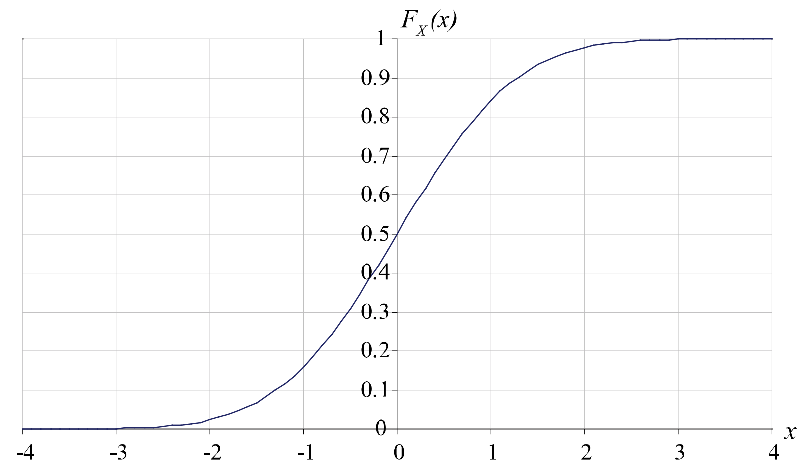
-> die Form eine geraden Linie ist z.B. gegeben durch:

$$x = a + by$$

Schätzung und Modellentwicklung

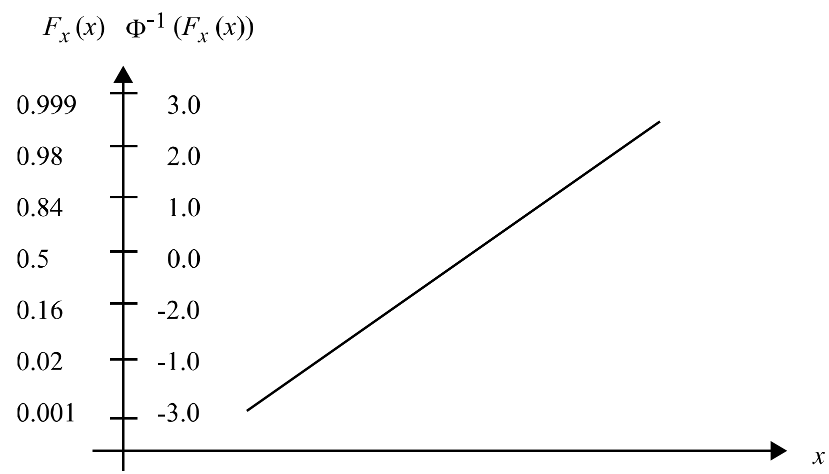
Modellauswahl anhand von Wahrscheinlichkeitspapier

Beispiel:
Wahrscheinlichkeitspapier für eine normalverteilte Wahrscheinlichkeits-verteilungsfunktion



$$F_X(x) = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right)$$

$$x = \Phi^{-1}(F_X(x)) \cdot \sigma_X + \mu_X$$



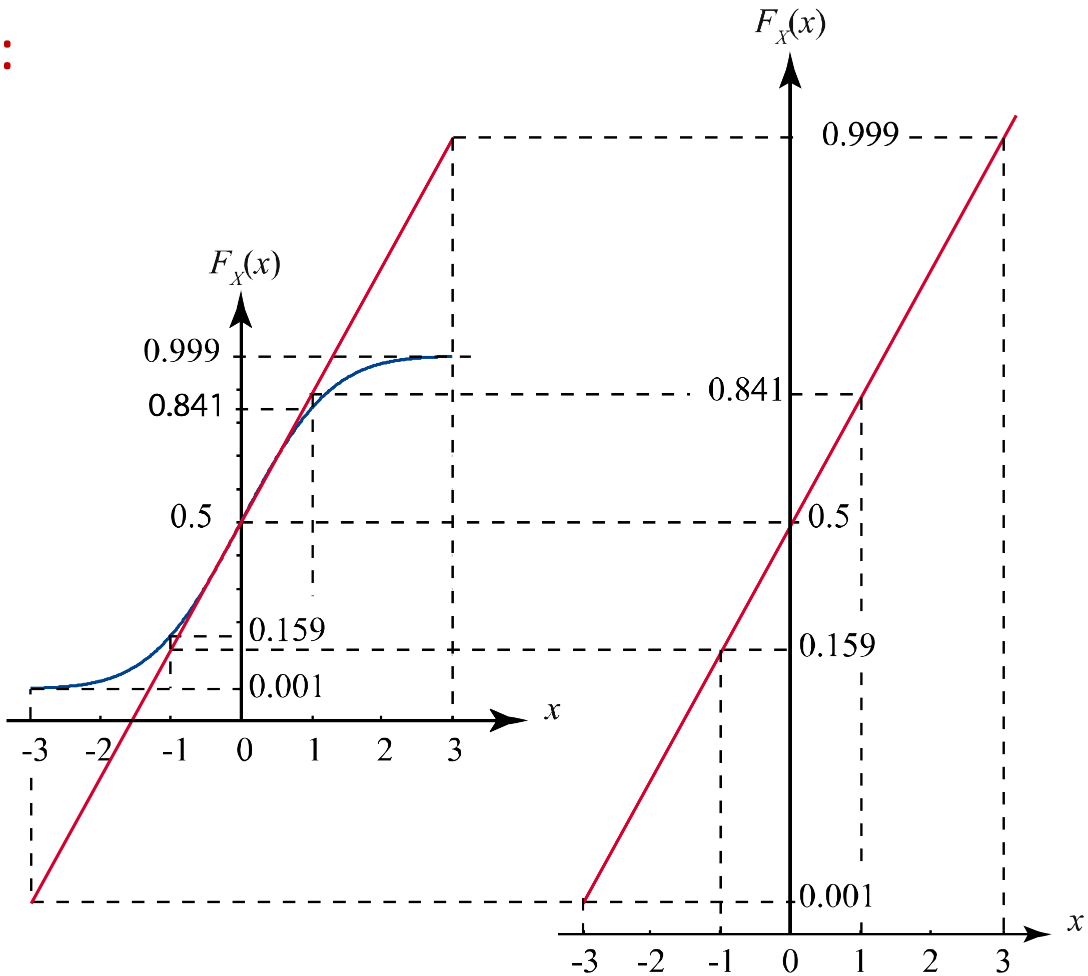
Die y-Achse ist nicht-linear skaliert.

Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Grafischer Ansatz:

Normalverteilung



Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Die Stichproben-Verteilungsfunktion kann anhand einer sortierten Messreihe abgeleitet werden:

$$F_X(x_i) = \frac{i}{N+1}$$

Beispiel: Druckfestigkeit von Beton

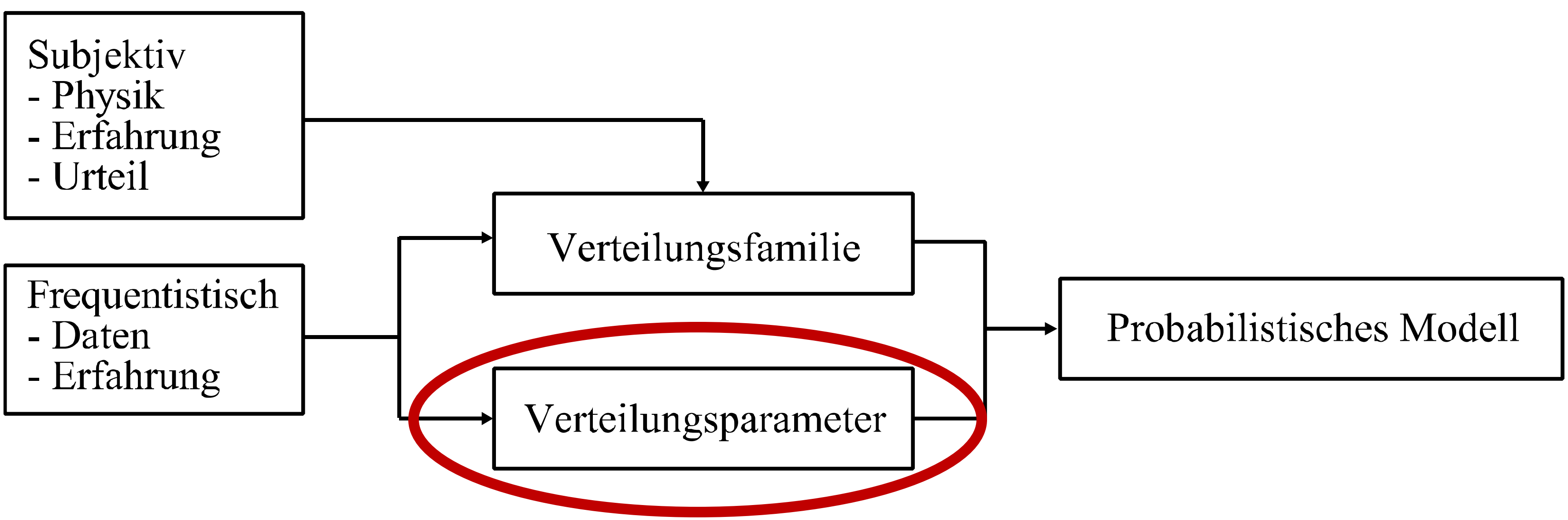
Lösung:
Normalverteilungs-Wahrscheinlichkeitspapier

i	$\hat{x}_i^o$	$F_X(\hat{x}_i^o)$	$\Phi^{-1}(F_X(\hat{x}_i^o))$
1	24.4	0.048	- 1.668
2	27.6	0.095	- 1.309
3	27.8	0.143	- 1.068
4	27.9	0.190	- 0.876
5	28.5	0.238	- 0.712
6	30.1	0.286	- 0.566
7	30.3	0.333	- 0.431
8	31.7	0.381	- 0.303
9	32.2	0.429	- 0.180
10	32.8	0.476	- 0.060
11	33.3	0.524	0.060
12	33.5	0.571	0.180
13	34.1	0.619	0.303
14	34.6	0.667	0.431
15	35.8	0.714	0.566
16	35.9	0.762	0.712
17	36.8	0.810	0.876
18	37.1	0.857	1.068
19	39.2	0.905	1.309
20	39.7	0.952	1.668

Schätzung und Modellentwicklung - Übersicht

Wenn man Modelle im Ingenieurbereich entwickeln möchte, müssen unterschiedliche Typen von Informationen herangezogen werden.

- subjektive Informationen
- frequentistische Informationen



Schätzung der Verteilungsparameter

Haben wir uns für ein Verteilungstyp entschieden, müssen die Parameter geschätzt werden.

z.B. Normalverteilung

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

Weibullverteilung

$$f_X(x) = \frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k-1} \exp\left(-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^k \right)$$

Schätzung der Verteilungsparameter

Haben wir uns für ein Verteilungstyp entschieden, müssen die Parameter geschätzt werden.

z.B. Normalverteilung

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

Weibullverteilung

$$f_X(x) = \frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k-1} \exp\left(-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k}\right)$$

Wahrscheinlichkeitsdichtefunktionen werden definiert durch ihre Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$.

Allgemein gibt man Dichtefunktionen bedingt auf die Parameter an: $f_X(x|\boldsymbol{\theta})$

Schätzung der Verteilungsparameter

Es gibt eine Vielzahl von Methoden, Verteilungsparameter zu schätzen; generell wird unterschieden zwischen:

- Punktschätzern
- Intervallschätzern.

Im Folgenden werden wir zwei Methoden näher betrachten:

- Methode der Momente
- Methode der Maximum Likelihood

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Gegeben ist eine Stichprobe anhand derer wir die Verteilungsparameter schätzen: $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_{2,}, .., \hat{x}_n)^T$

Das Prinzip der Methode der Momenten ist, die Parameter so abzuschätzen, dass die Momente der Verteilung und die Momente der Stichprobe identisch sind.

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Gegeben ist eine Stichprobe anhand derer wir die Verteilungsparameter schätzen: $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_{2,}, .., \hat{x}_n)^T$

Das Prinzip der Methode der Momenten ist, die Parameter so abzuschätzen, dass die Momente der Verteilung und die Momente der Stichprobe identisch sind.

$$m_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$$

Stichprobe

$$\lambda_j = \lambda_j(\theta_1, \theta_2, .., \theta_k) = \int_{-\infty}^{\infty} x^j \cdot f_X(x|\boldsymbol{\theta})dx$$

Verteilung

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Angenommen, wir haben eine Verteilung mit k Parametern:

Dann müssen wir k Gleichungen mit k Unbekannten lösen:

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Angenommen, wir haben eine Verteilung mit k Parametern:

Dann müssen wir k Gleichungen mit k Unbekannten lösen:

$$m_j = \lambda_j(\boldsymbol{\theta}),\ j = 1, 2, .., k$$

$$\Downarrow$$

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i^j} = \boxed{\int_{-\infty}^{\infty} x^j \cdot f_X(x|\boldsymbol{\theta})dx}, \qquad j = 1, 2, .., k$$

Stichprobe | Verteilung

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Beispiel:
Druckfestigkeit von Beton, Annahme einer Normalverteilung:

Die Normalverteilung hat zwei Parameter – wir müssen also zwei Gleichungen lösen:

$$\boxed{m_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i} = \boxed{\lambda_1 = \int_{-\infty}^{\infty} x \cdot f_X(x|\mu,\sigma)dx}$$

$$\boxed{m_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2} = \boxed{\lambda_2 = \int_{-\infty}^{\infty} x^2 \cdot f_X(x|\mu,\sigma)dx}$$

Betondruck-festigkeit [Mpa]
24.4
27.6
27.8
27.9
28.5
30.1
30.3
31.7
32.2
32.8
33.3
33.5
34.1
34.6
35.8
35.9
36.8
37.1
39.2
39.7

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Die Stichprobenmomente sind:

$$m_1 = \frac{1}{20}\sum_{i=1}^{n} \hat{x}_i = 32.67 \qquad m_2 = \frac{1}{20}\sum_{i=1}^{n} \hat{x}_i^2 = 1083.36$$

Die Momente der Verteilung sind:

$$\lambda_1 = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-0.5\frac{(x-\mu)^2}{\sigma^2})dx \qquad \lambda_2 = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp(-0.5\frac{(x-\mu)^2}{\sigma^2})dx$$

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Bei Formulierung der folgenden Objektfunktion:

$$g(\mu,\sigma) = (\lambda_1(\mu,\sigma) - m_1)^2 + (\lambda_2(\mu,\sigma) - m_2)^2$$

Lassen sich die Parameter durch Optimierung finden.

z.B. mit Hilfe von MS EXCEL.

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

ODER:

Alternativ und viel einfacher ist die Herangehensweise mit Hilfe der **zentralen** Momente:

1. Bestimmung des Stichprobenmittelwertes und der Stichprobenstandardabweichung.
2. Gleichsetzen mit dem Mittelwert und der Standardabweichung der Verteilungsfunktion.
3. Berechnung der Parameter.

Schätzung der Verteilungsparameter

- Methode der Momente (MoM)

Log-Normalverteilung,	Parameter	Erwartungswert und Streuung
$F_X(x) = \Phi\left(\frac{\ln(x)-\lambda}{\zeta}\right)$ $f_X(x) = \frac{1}{x\zeta\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\zeta}\right)^2\right)$	$0 \le x \le \infty$ λ, ζ $\zeta > 0$	$\mu = \exp\left(\lambda + \frac{\zeta^2}{2}\right)$ $\sigma = \mu\sqrt{\exp(\zeta^2)-1}$
Gumbel max.	**Parameter**	**Erwartungswert und Streuung**
$f_X(x) = \alpha\exp(-\alpha(x-u)-\exp(x-u))$ $F_X(x) = \exp(-\exp(-\alpha(x-u)))$	$-\infty \le x \le \infty$ u $\alpha > 0$	$\mu = u + \frac{\gamma}{\alpha}$, $\gamma \approx 0.5772$ $\sigma = \frac{\pi}{\alpha\sqrt{6}}$

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

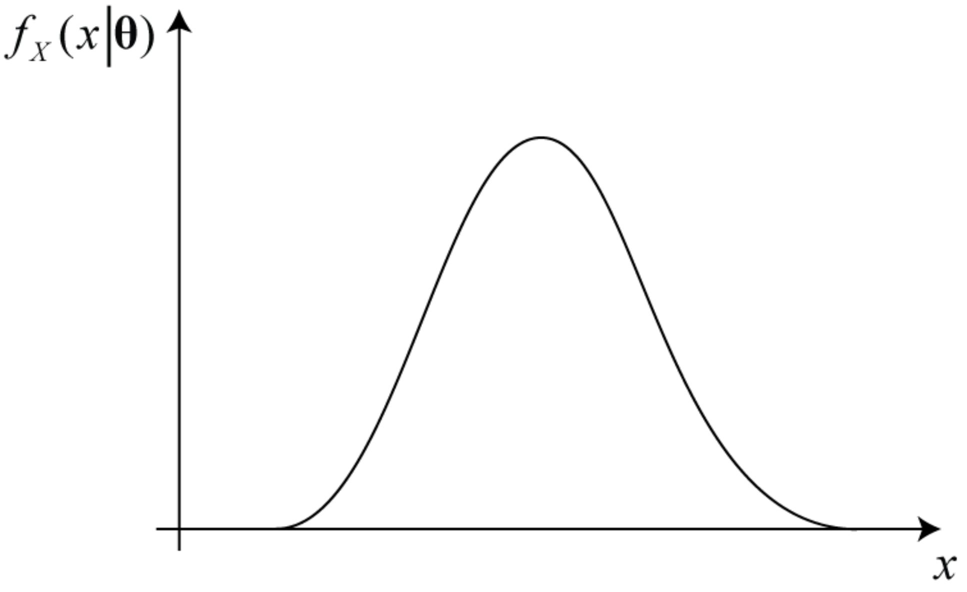
Die Grundidee der MLM ist:

die Parameter der Verteilung zu identifizieren, welche die maximale ‚Likelihood' besitzen – welche also am wahrscheinlichsten sind.

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Die Form einer Wahrscheinlichkeitsdichtefunktion ist bestimmt durch ihre Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$:



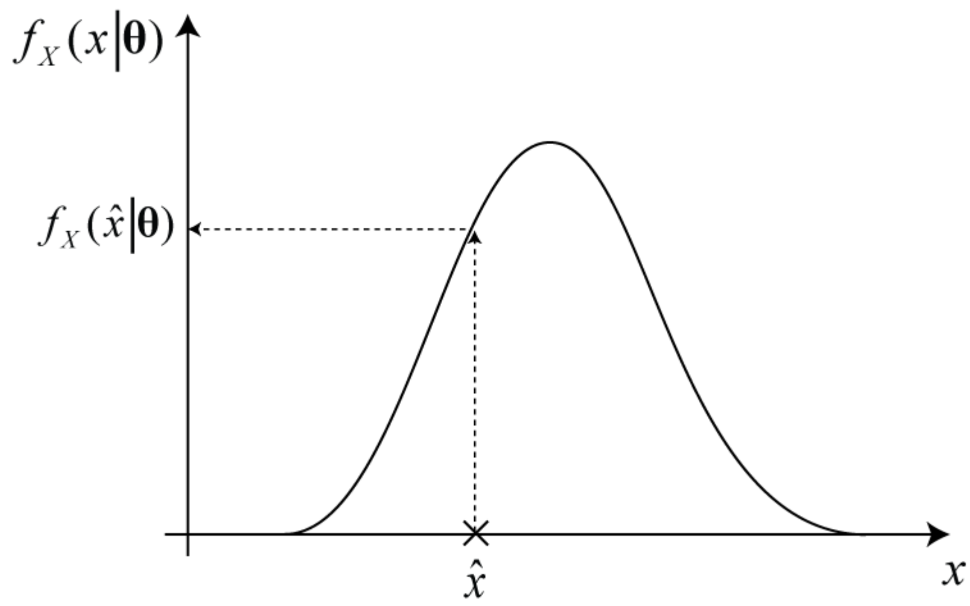
Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Die Form einer Wahrscheinlichkeitsdichtefunktion ist bestimmt durch ihre Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$.

Wird nun eine Stichprobe $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, .., \hat{x}_n)^T$ beobachtet, hat eine Realisation der Stichprobe $\hat{x}$ die ‚Likelihood':

$$L = f_X(\hat{x} | \boldsymbol{\theta})$$



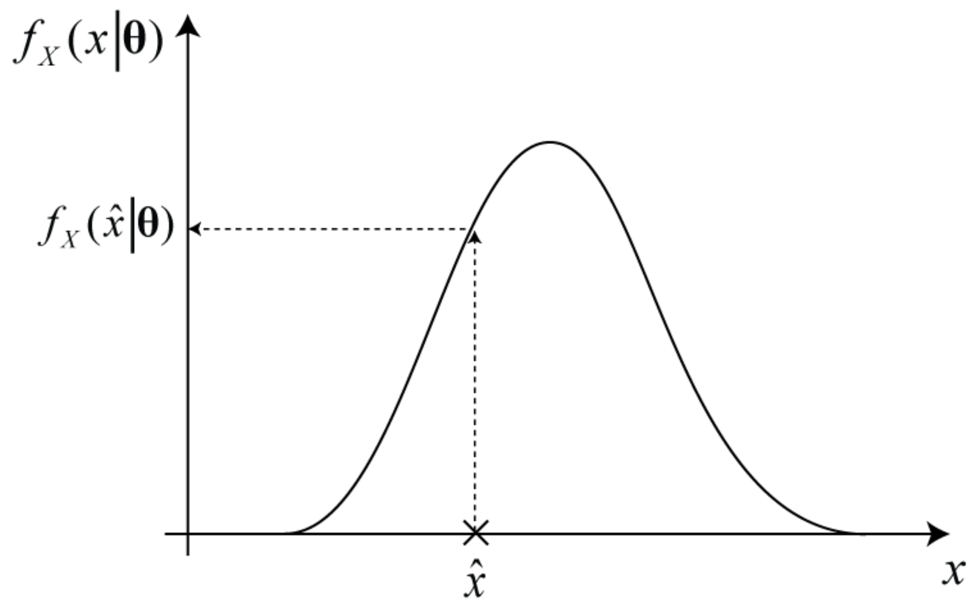
Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Die Form einer Wahrscheinlichkeitsdichtefunktion ist bestimmt durch ihre Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$:

Wird nun eine Stichprobe $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, .., \hat{x}_n)^T$ beobachtet, hat eine Realisation der Stichprobe $\hat{x}$ die ‚Likelihood':

$$L = f_X(\hat{x} | \boldsymbol{\theta})$$



Die ‚Likelihood' der gesamten Stichproben errechnet sich aus dem Produkt der Likelihoods der Einzelbeobachtungen -> $L = \prod_{i=1}^{n} f_X(\hat{x}_i | \boldsymbol{\theta})$

→ bei gegebenen Parametern

Schätzung der Verteilungsparameter

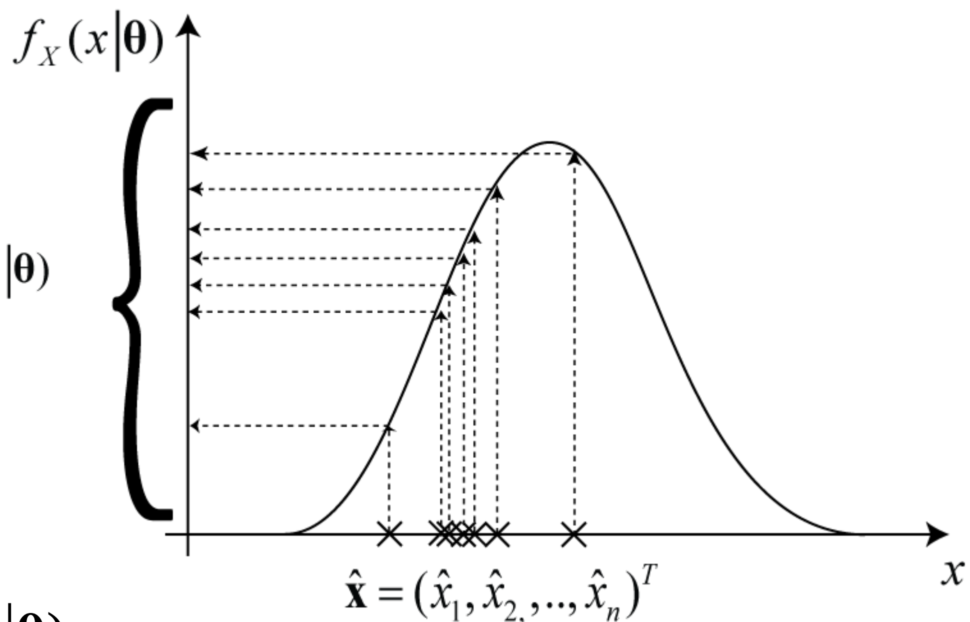
- Maximum Likelihood Methode (MLM)

Die Form einer Wahrscheinlichkeitsdichtefunktion ist bestimmt durch ihre Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$.

Wird nun eine Stichprobe $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, .., \hat{x}_n)^T$ beobachtet, hat eine Realisation der Stichprobe $\hat{x}$ die ‚Likelihood':

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

$$L = f_X(\hat{x} | \boldsymbol{\theta})$$



Die ‚Likelihood' der gesamten Stichproben errechnet sich aus dem Produkt der Likelihoods der Einzelbeobachtungen -> $L = \prod_{i=1}^{n} f_X(\hat{x}_i | \boldsymbol{\theta})$

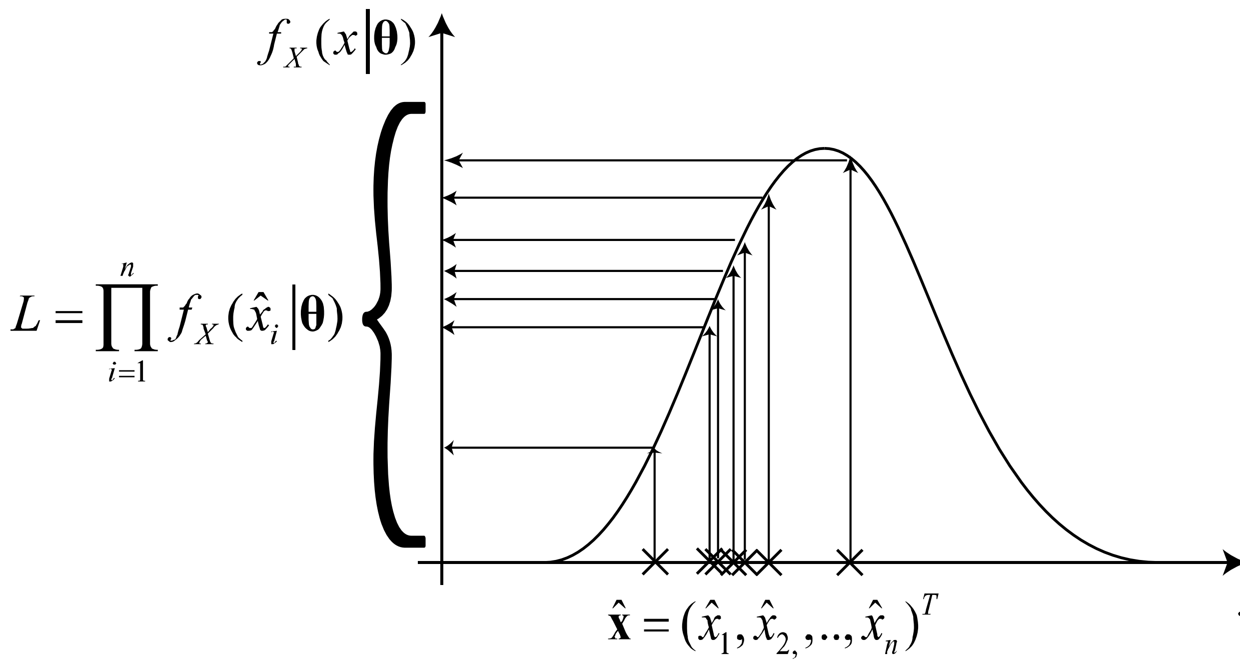
→ bei gegebenen Parametern

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Wir maximieren nun die ‚Likelihood' unter Veränderung der Parameter:

$$\max_{\boldsymbol{\theta}} L\left(\boldsymbol{\theta} \middle| \hat{x}_i\right)$$



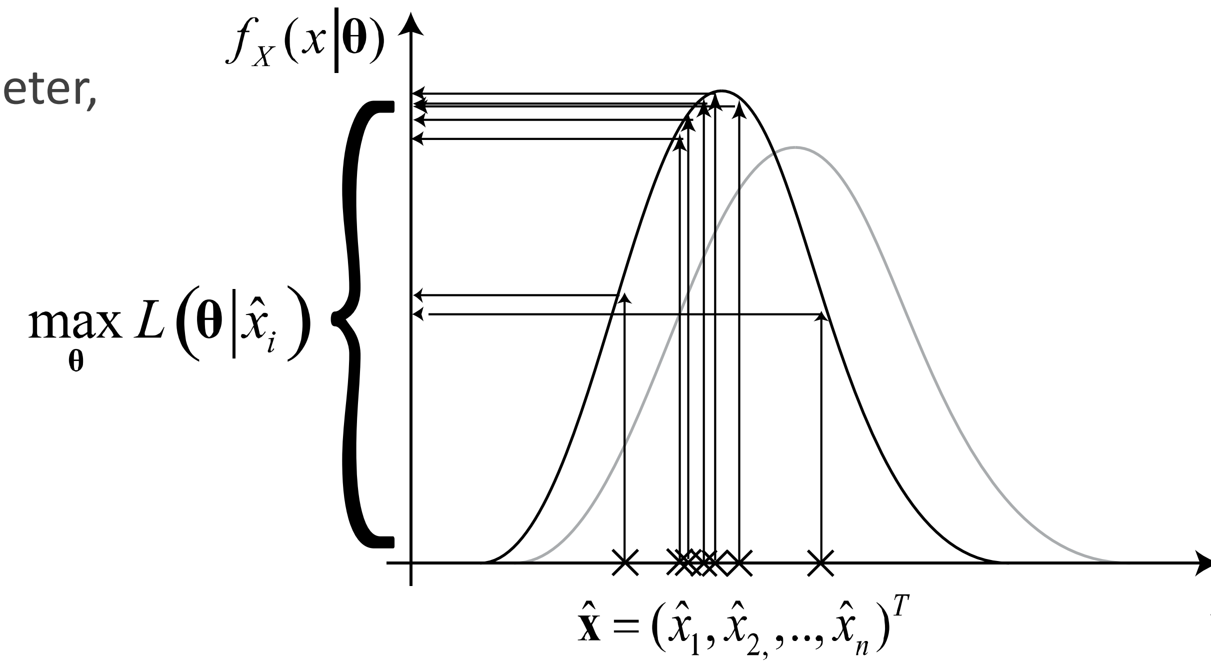
Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Wir maximieren nun die ‚Likelihood' unter Veränderung der Parameter:

$$\max_{\boldsymbol{\theta}} L\left(\boldsymbol{\theta} \middle| \hat{x}_i\right)$$

→ und erhalten die Parameter, welche am besten zu der Stichprobe passen.



Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Wir nehmen an, die unseren Beobachtungen zugrundeliegende Verteilungsfunktion sei die Normalverteilung.

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

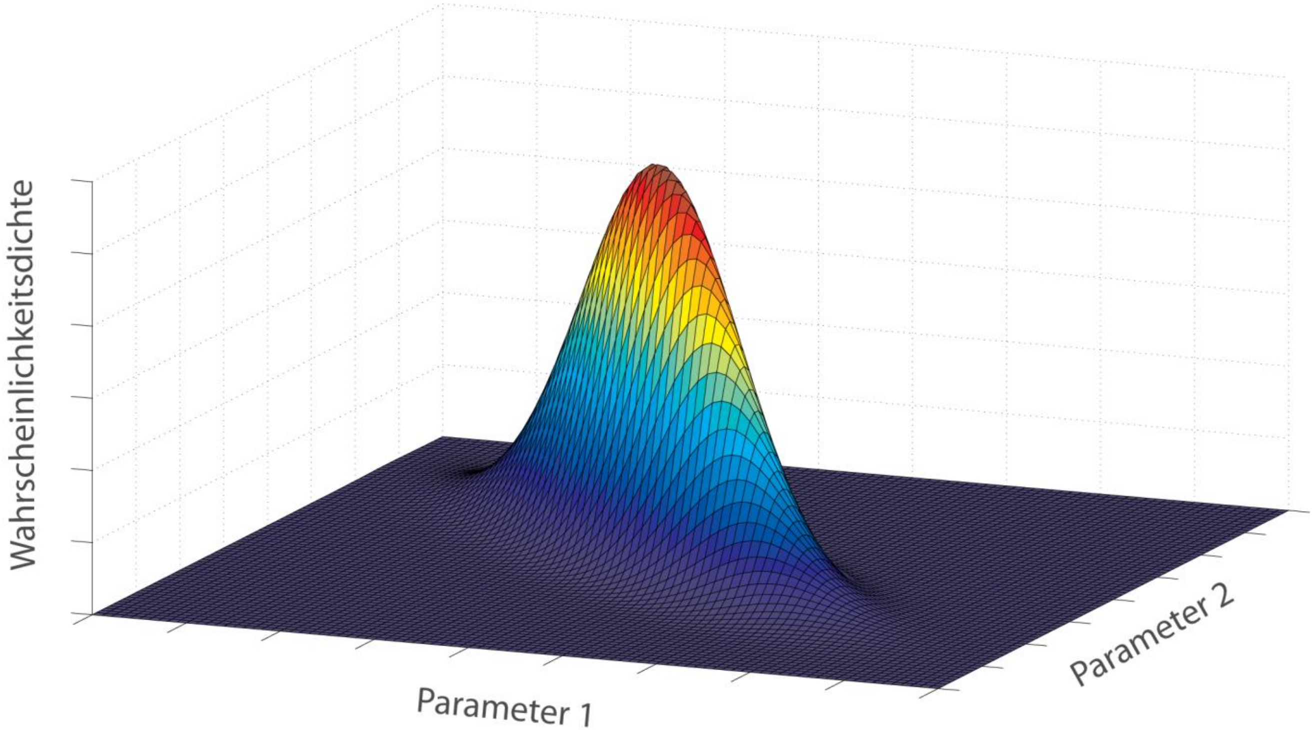
Die Likelihood der Beobachtungen der Stichprobe $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)^T$ ist dann;

$$L(\boldsymbol{\theta}|\hat{\mathbf{x}}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\hat{x}_i-\mu}{\sigma}\right)^2\right)$$

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Illustration



Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Die Parameter der Normalverteilung $\boldsymbol{\theta}$ werden so gewählt, dass sie die Likelihoodfunktion maximieren:

$$\min_{\boldsymbol{\theta}}(-L(\boldsymbol{\theta}|\hat{\mathbf{x}}))$$

Es ist von Vorteil, die logarithmierte Likelihoodfunktion zu verwenden:

$$l(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \ln(f_X(\hat{x}_i|\boldsymbol{\theta})) \qquad -l(\boldsymbol{\theta}|\mathbf{x}) = -\sum_{i=1}^{n} \ln(f_X(\hat{x}_i|\boldsymbol{\theta}))$$

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Wir wählen die Parameter $\boldsymbol{\theta}$ so, dass sie die log-Likelihoodfunktion maximieren:

$$\min_{\boldsymbol{\theta}}(-l(\boldsymbol{\theta}|\hat{\mathbf{x}}))$$

Es kann gezeigt werden, dass die Parameter selbst zu normalverteilten Zufallsvariablen konvergieren:

Mit Mittelwerten: $\boldsymbol{\mu}_{\Theta} = (\theta_1^*, \theta_2^*, .., \theta_n^*)^T$

Und Kovarianzmatrix: $\mathbf{C}_{\Theta\Theta} = \mathbf{H}^{-1}$ wobei $H_{ij} = \left.\dfrac{\partial^2 - l(\boldsymbol{\theta}|\hat{\mathbf{x}})}{\partial\theta_i \partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Betrachten wir nun die Daten der Betondruckfestigkeit. Unter der Annahme, dass die Betondruckfestigkeit einer Normalverteilung folgt, ergibt sich für die log-Likelihoodfunktion:

$$l(\boldsymbol{\theta}|\hat{\mathbf{x}}) = n \cdot \ln\left(\frac{1}{\sqrt{2\pi}\theta_1}\right) - \frac{1}{2}\sum_{i=1}^{n}\frac{(\hat{x}_i - \theta_2)^2}{\theta_1^2}$$

Das Minimum kann analytisch wie folgt bestimmt werden:

$$\frac{\partial l}{\partial \theta_1} = -\frac{n}{\theta_1} + \frac{1}{\theta_1^3}\sum_{i=1}^{n}(\hat{x}_i - \theta_2)^2 = 0$$

$$\frac{\partial l}{\partial \theta_2} = \frac{1}{\theta_1^2}\sum_{i=1}^{n}(\hat{x}_i - \theta_2) = 0$$

$$\Rightarrow$$

$$\theta_1 = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - \theta_2)^2}{n}}$$

$$\theta_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i$$

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Gemäss Zahlenbeispiel ‚Betondruckfestigkeit' ergibt sich:

$$\theta_1 = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{x}_i - \theta_2\right)^2}{n}} = \sqrt{\frac{367.19}{20}} = 4.05$$

Mittelwert der Standardabweichung – nicht erwartungstreu (!)

$$\theta_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i = \frac{653.3}{20} = 32.67$$

Mittelwert des Mittelwertes

Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Beispiel Normalverteilung:
Gemäss Zahlenbeispiel ‚Betondruckfestigkeit' ergibt sich:

Für die Kovarianzmatrix:

$$H = \begin{pmatrix} \dfrac{n}{\theta_1} - \dfrac{3\sum_{i=1}^{n}(x_i - \theta_2)^2}{\theta_1^4} & \dfrac{2\sum_{i=1}^{n}(x_i - \theta_2)}{\theta_1^3} \\ \dfrac{2\sum_{i=1}^{n}(x_i - \theta_2)}{\theta_1^3} & \dfrac{n}{\theta_1^2} \end{pmatrix}$$

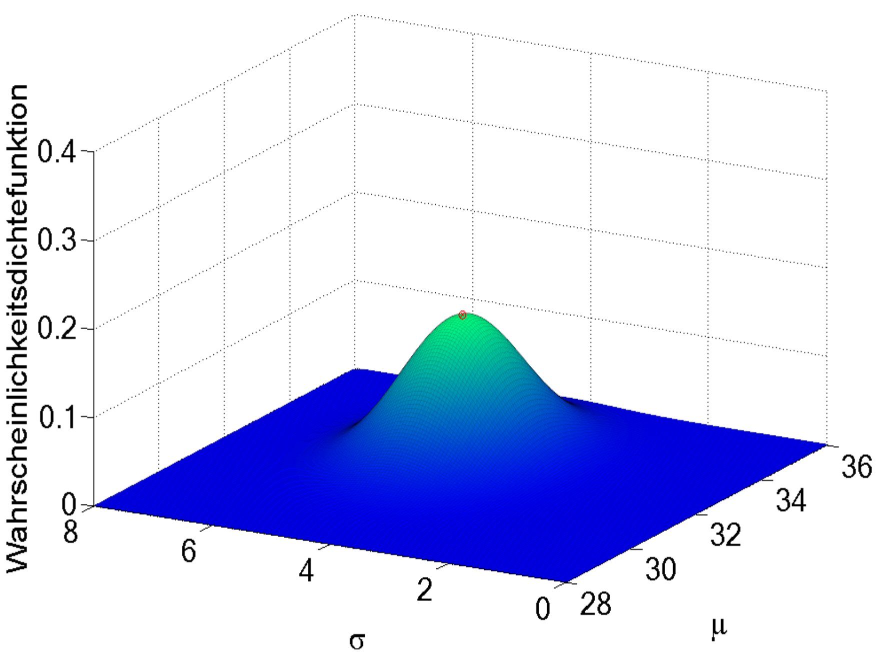
$$C_{\Theta\Theta} = H^{-1} = \begin{pmatrix} \boxed{0.836} & 0 \\ 0 & \boxed{0.165} \end{pmatrix}$$

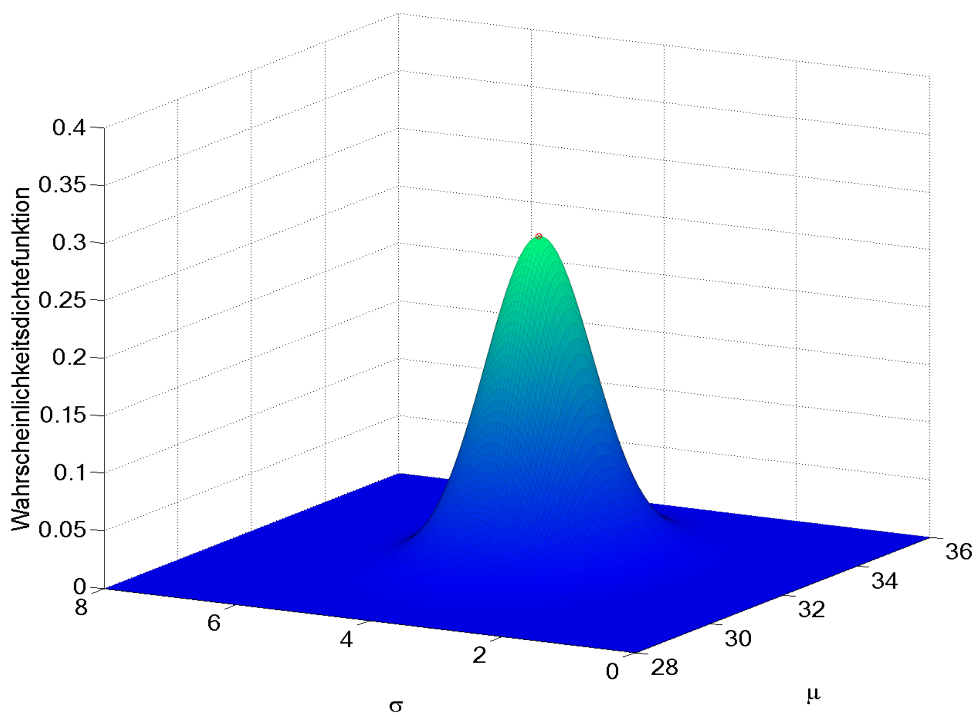
Varianz des Mittelwertes

Varianz der Standardabweichung

Schätzung der Verteilungsparameter

Mit Hilfe der Fisher-Informationsmatrix können die Unsicherheiten in der Schätzung der Parameter bestimmt werden.



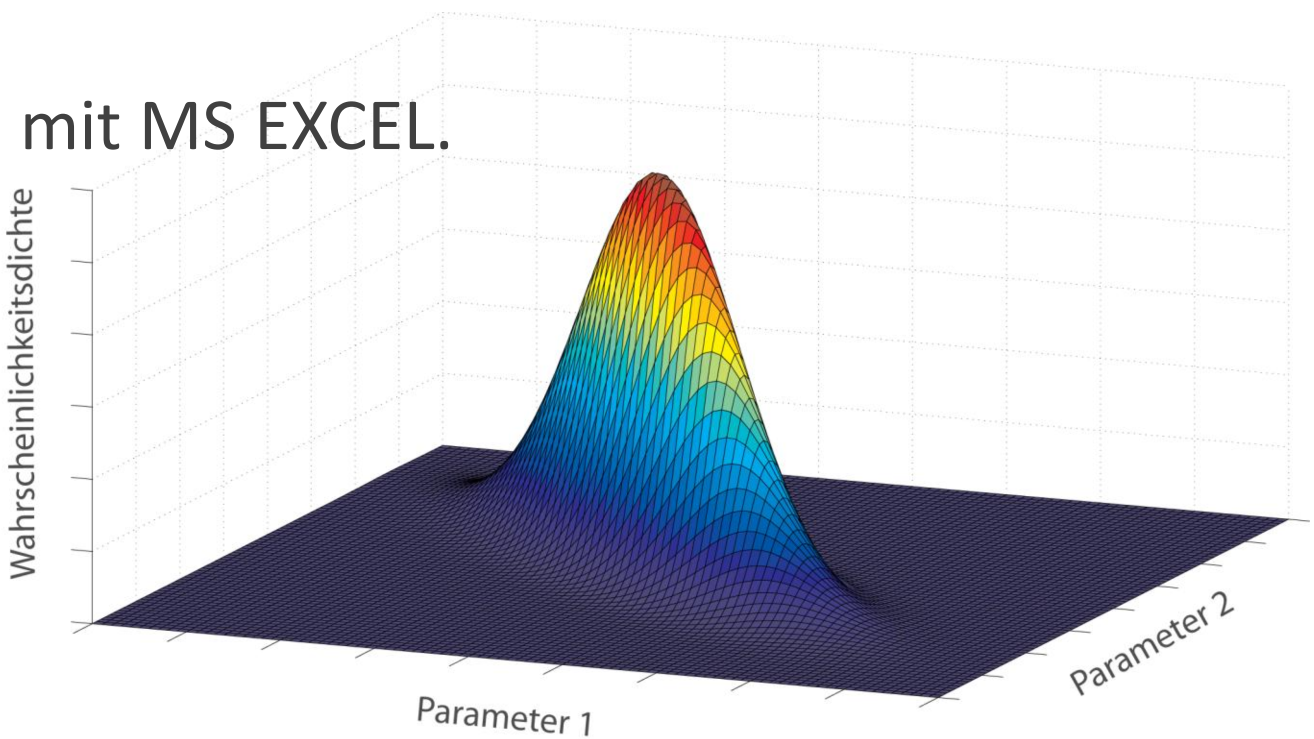


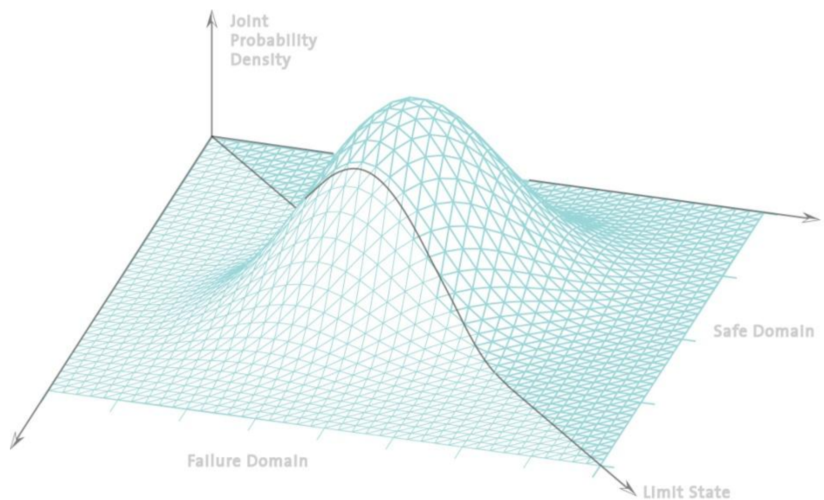
Schätzung der Verteilungsparameter

- Maximum Likelihood Methode (MLM)

Das Problem lässt sich ebenfalls numerisch lösen

– z.B. mit MS EXCEL.





Statistik und Wahrscheinlichkeitsrechnung

Prof. Dr. Michael Havbro Faber