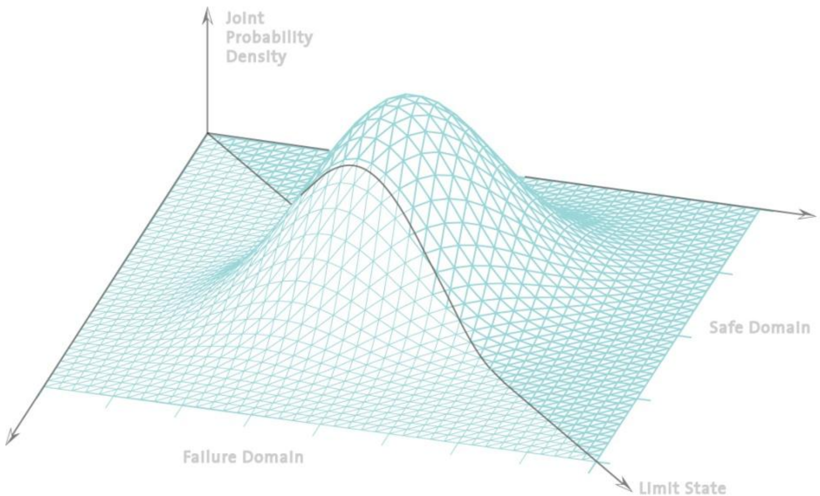


Statistik und Wahrscheinlichkeitsrechnung

Übung 8

Inhalt der heutigen Übung

Aufgabe E.5: Wahrscheinlichkeitspapier

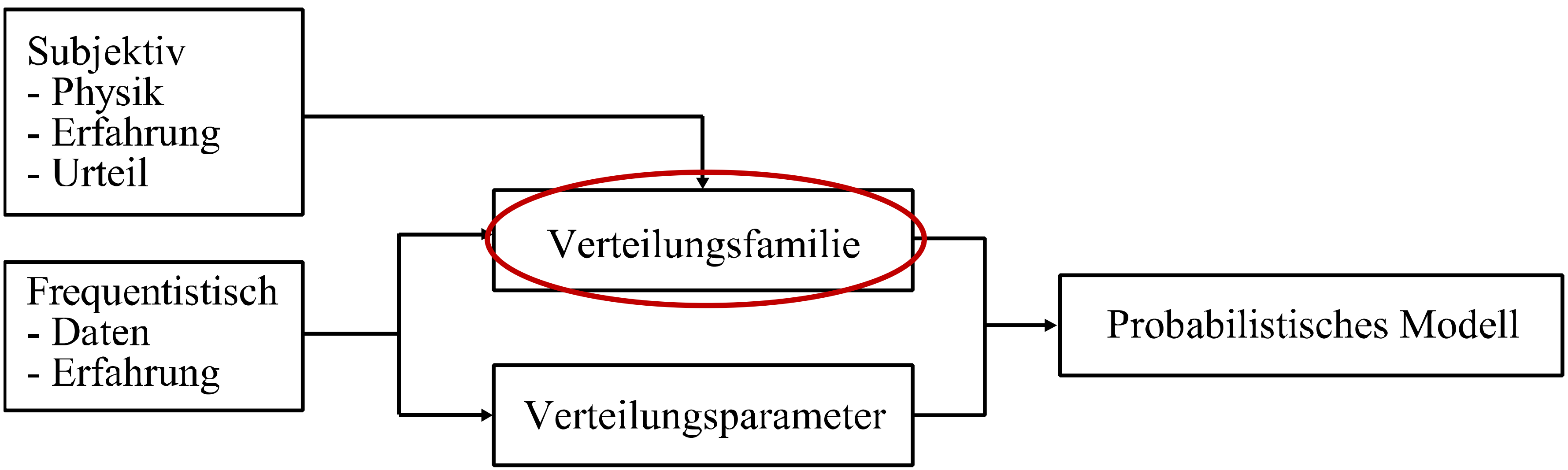
Aufgabe E.8: Maximum-Likelihood-Methode

Aufgaben E.11:Methode der Momente

Hausübung E.13

Wahrscheinlichkeitspapier

Problemstellung:



Wahrscheinlichkeitspapier

Wir wollen wissen, ob eine Stichprobe mit einer bestimmten Verteilungsfamilie beschrieben werden kann.

Hierfür kann für eine erste Abschätzung ohne Bestimmung der Verteilungs-parameter eine qualitative Überprüfung mit Hilfe von einem Wahrscheinlichkeitspapier durchgeführt werden.

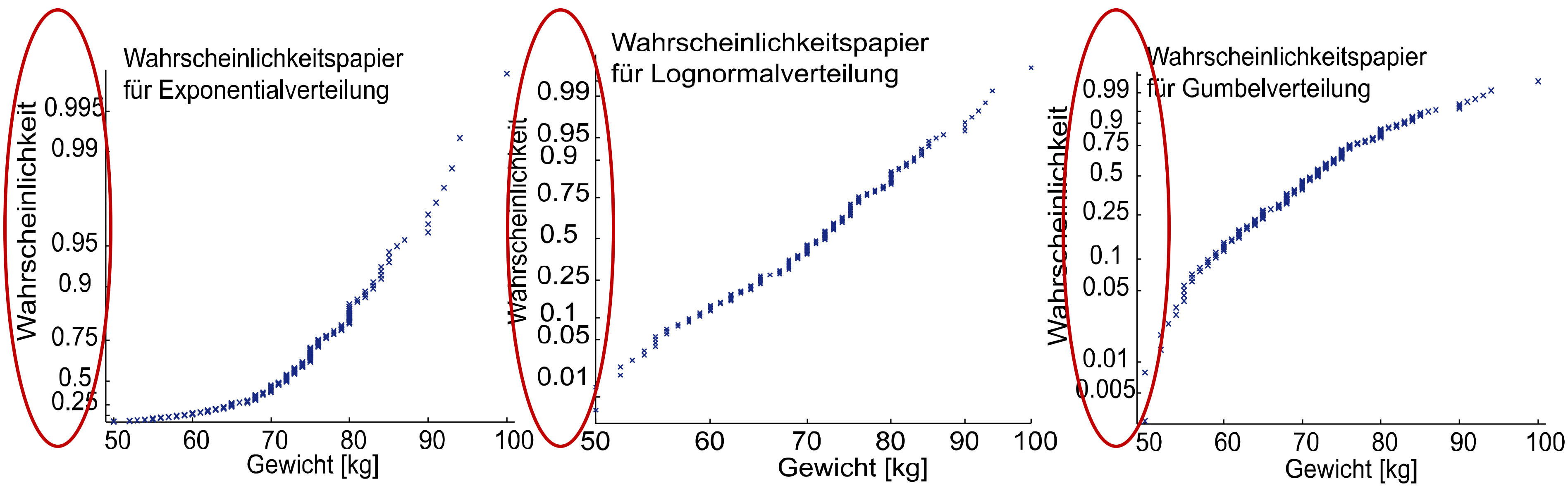
- ✓ Wenn die Punkte der beobachteten Werte auf dem Wahrscheinlichkeitspapier in etwa auf einer Geraden liegen, dann kann die Verteilungsfamilie <u>qualitativ</u> akzeptiert werden.

Wie kann eine Punktmenge auf einer Linie liegen?

→ Linearisierung der y-Achse für die betrachtete Verteilungsfamilie

Wahrscheinlichkeitspapier

Ein Beispiel: Zur Beschreibung des Körpergewichtes der Studierenden aus der Statistik-Vorlesung wird eine geeignete Verteilungsfamilie gesucht.



→ Das Körpergewicht der Studierenden folgt am ehesten einer Lognormalverteilung.

Aufgabe E.5

Aus Verkehrszählungen liegt eine Datenserie vor, die den täglichen Verkehrsfluss in der Rosengartenstrasse in Zürich beschreibt.

a) Erstelle ein Wahrscheinlichkeitspapier für folgende Dichtefunktion

$$f_X(x) = \begin{cases} \dfrac{2}{10000^2}\,x & 0 \le x \le 10000 \\ 0 & \text{sonst} \end{cases}$$

b) Überprüfe mit Hilfe des erstellten Wahrscheinlichkeitspapiers, ob der tägliche Verkehrsfluss mit dieser Verteilung beschrieben werden kann.

Beobachtung	Anzahl Fahrzeuge
1	3600
2	4500
3	5400
4	6500
5	7000
6	7500
7	8700
8	9000
9	9500

Aufgabe E.5 – Lösung

Wahrscheinlichkeits-dichtefunktion

$$f_X(x) = \begin{cases} \dfrac{2}{10000^2}x & 0 \le x \le 10000 \\ 0 & \text{sonst} \end{cases}$$

Wahrscheinlichkeits-verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & 0 \le x \\ \left(\dfrac{x}{10000}\right)^2 & 0 < x \le 10000 \\ 1 & x > 10000 \end{cases}$$

Linearisierung

$$F_X(x) = \left(\frac{x}{10000}\right)^2 \Leftrightarrow \sqrt{F_X(x)} = \frac{x}{10000}$$

Aufgabe E.5 – Lösung

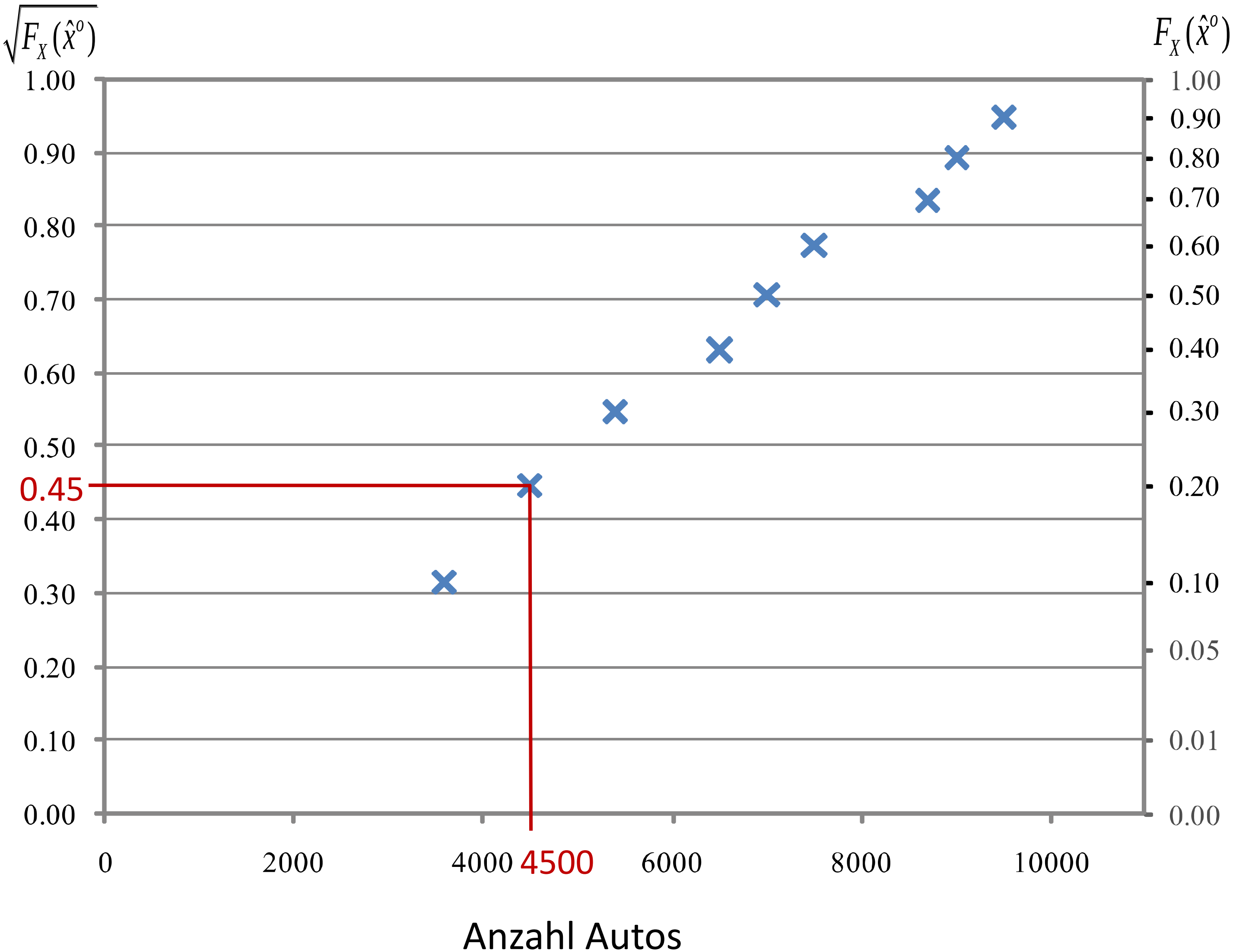
Eine linearisierte y-Achse erhalten wir, indem wir für alle Quantile in bestimmten Abständen die linearisierte Form berechnen - beispielsweise zu jedem Datenpunkt.

Berechnung des Quantilindex:

$$\upsilon = \frac{i}{n+1} = F_X(\hat{x}^o)$$

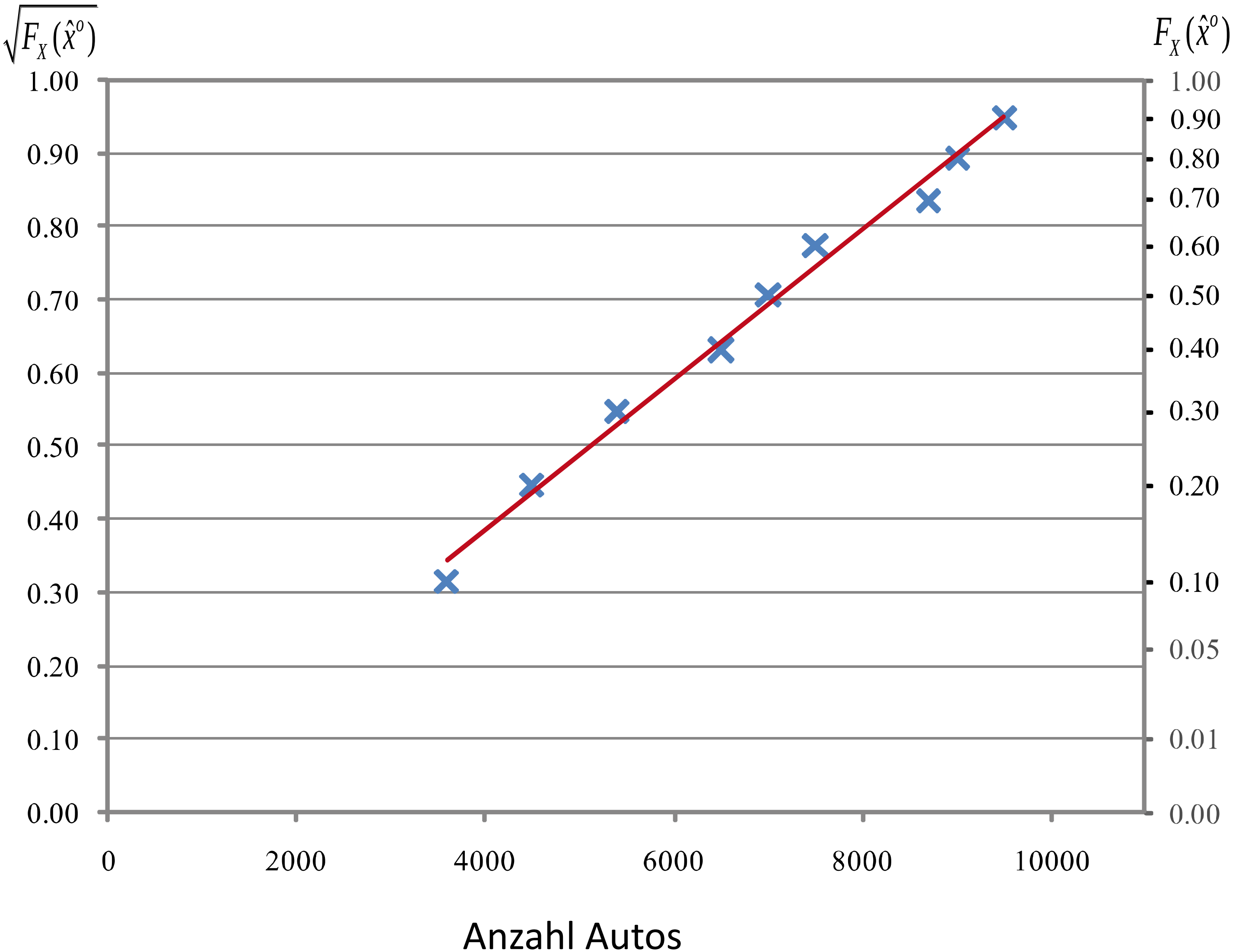
Rang i	Anzahl Autos (x-Achse)	$F_X(\hat{x}^o)$	$\sqrt{F_X(\hat{x}^o)}$
		0	0
1	3600	0.1	0.32
2	4500	0.2	0.45
3	5400	0.3	0.55
4	6500	0.4	0.63
5	7000	0.5	0.71
6	7500	0.6	0.77
7	8700	0.7	0.84
8	9000	0.8	0.89
9	9500	0.9	0.95
		1.0	1.0

Aufgabe E.5 – Lösung



Anzahl Autos	$F_X(\hat{x}^0)$	$\sqrt{F_X(\hat{x}^0)}$
	0	0
3600	0.1	0.32
4500	0.2	0.45
5400	0.3	0.55
6500	0.4	0.63
7000	0.5	0.71
7500	0.6	0.77
8700	0.7	0.84
9000	0.8	0.89
9500	0.9	0.95
	1.0	1.0

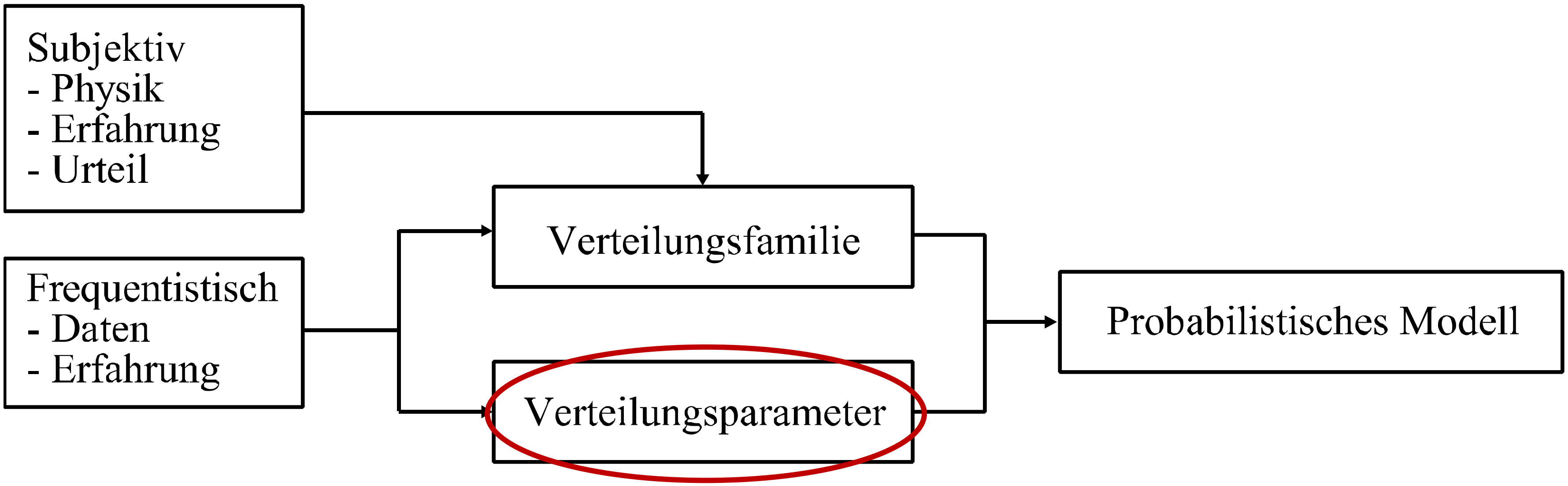
Aufgabe E.5 – Lösung



Anzahl Autos	$F_X(\hat{x}^0)$	$\sqrt{F_X(\hat{x}^0)}$
	0	0
3600	0.1	0.32
4500	0.2	0.45
5400	0.3	0.55
6500	0.4	0.63
7000	0.5	0.71
7500	0.6	0.77
8700	0.7	0.84
9000	0.8	0.89
9500	0.9	0.95
	1.0	1.0

Schätzung der Verteilungsparameter

- Methode der Momente
- Maximum Likelihood Methode



Maximum-Likelihood-Methode (MLM)

multivariate Dichtefunktion der Beobachtungen mit gegebenen Parametern:

X_i unabhängig   gleiche Verteilung der X_i

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_i}(x_i \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{\theta})$$

Wir können in diesem Ausdruck die Beobachtungen $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ einsetzen...

... und erhalten dadurch die **Likelihood-Funktion**:

$$L(\boldsymbol{\theta} \mid \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \prod_{i=1}^{n} f_X(\hat{x}_i \mid \boldsymbol{\theta})$$

Maximum-Likelihood-Methode (MLM)

Die Maximum-Likelihood-Schätzwerte von $\boldsymbol{\theta}$ erhält man, indem man die Werte von $\boldsymbol{\theta}$ wählt, die die Likelihood-Funktion $L(\boldsymbol{\theta})$ maximieren.

$$L(\boldsymbol{\theta} \mid \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \prod_{i=1}^{n} f_X(\hat{x}_i \mid \boldsymbol{\theta})$$

Dies ist gleichbedeutend mit der Maximierung der „Log-Likelihood-Funktion", die wie folgt definiert ist:

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

Aufgabe E.8

Um die Druckfestigkeit von Beton einer bestimmten Produktionsmarge zu modellieren, wurden 20 Stichproben gemessen. Es wird angenommen, dass die Grundgesamtheit der Stichproben einer Normalverteilung folgt.

a) Beschreibe die Likelihood-Funktion.

b) Schätze die unbekannten Parameter (μ,σ) anhand der MLM.

Was passiert, wenn die Normalverteilung durch eine Exponentialverteilung ersetzt wird?

c) Schätze die Parameter der Exponential-verteilung mit der MLM und vergleiche die kumulative Verteilungsfunktion mit den beobachteten Werten.

Nummer der Messung	Druck-festigkeit [MPa]	Nummer der Messung	Druck-festigkeit [MPa]
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

Aufgabe E.8

a) Die Likelihood-Funktion ist: $$L(\boldsymbol{\theta} \mid \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) = \prod_{i=1}^{n} f_X(\hat{x}_i \mid \boldsymbol{\theta})$$

für die Normalverteilung: $$L(\mu, \sigma \mid \hat{\mathbf{x}}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\hat{x}_i - \mu}{\sigma}\right)^2\right]$$

… und muss nun maximiert werden. Anstelle der Likelihood-Funktion empfiehlt es sich, die Log-Likelihood-Funktion zu verwenden:

$$l = \ln L = \sum_{i=1}^{n} \ln\left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\hat{x}_i - \mu}{\sigma}\right)^2\right]\right]$$

$$= -n\ln(\sqrt{2\pi}) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\hat{x}_i - \mu)^2$$

Aufgabe E.8

b) Die ML-Schätzwerte sind die Funktionsparameter, die die Likelihood-Funktion maximieren.

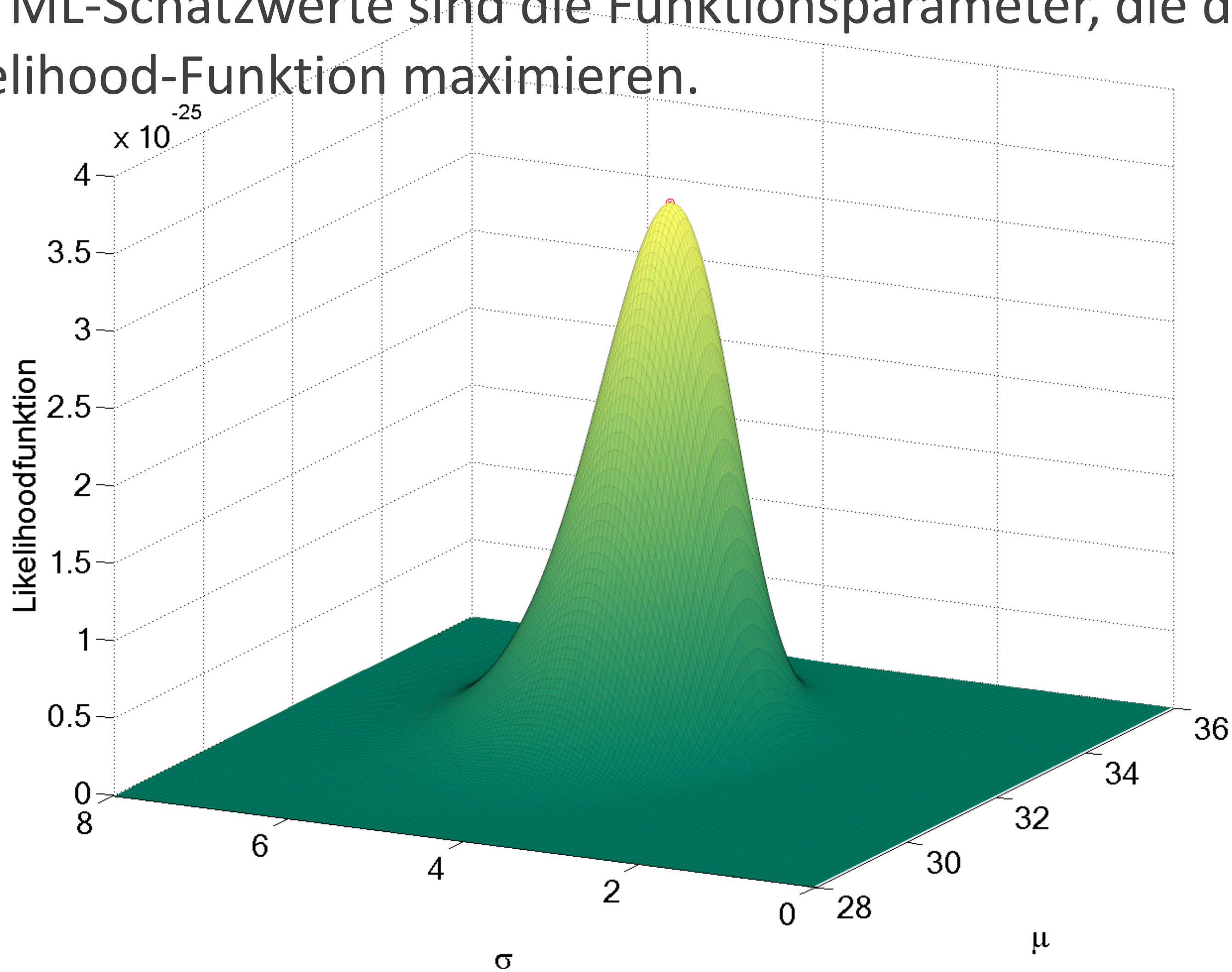
$$l = \ln L = \sum_{i=1}^{n} \ln\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\hat{x}_i-\mu)^2}{2\sigma^2}\right]\right]$$

$$= -n\ln(\sqrt{2\pi}) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\hat{x}_i-\mu)^2$$

$$\left.\begin{aligned}\frac{\partial l}{\partial \mu} &= \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(\hat{x}_i-\mu) = 0\\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(\hat{x}_i-\mu)^2 = 0\end{aligned}\right\} \quad \begin{aligned}\mu &= \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\\ \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i-\mu)^2\end{aligned} \quad \rightarrow \quad \begin{aligned}\mu &= 32.67\\ \sigma &= 4.04\end{aligned}$$

Aufgabe E.8

b) Die ML-Schätzwerte sind die Funktionsparameter, die die Likelihood-Funktion maximieren.



Fisher-Informationsmatrix

Mit Hilfe der Fisher-Informationsmatrix lässt sich die Variabilität und die Kovarianz der geschätzten Verteilungsparameter aus der Maximum-Likelihood-Methode ermitteln.

$$H = \begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{2\sum_{i=1}^{n}(\hat{x}_i - \mu)}{\sigma^3} \\ \dfrac{2\sum_{i=1}^{n}(\hat{x}_i - \mu)}{\sigma^3} & -\dfrac{n}{\sigma^2} + \dfrac{3\sum_{i=1}^{n}(\hat{x}_i - \mu)^2}{\sigma^4} \end{pmatrix}$$

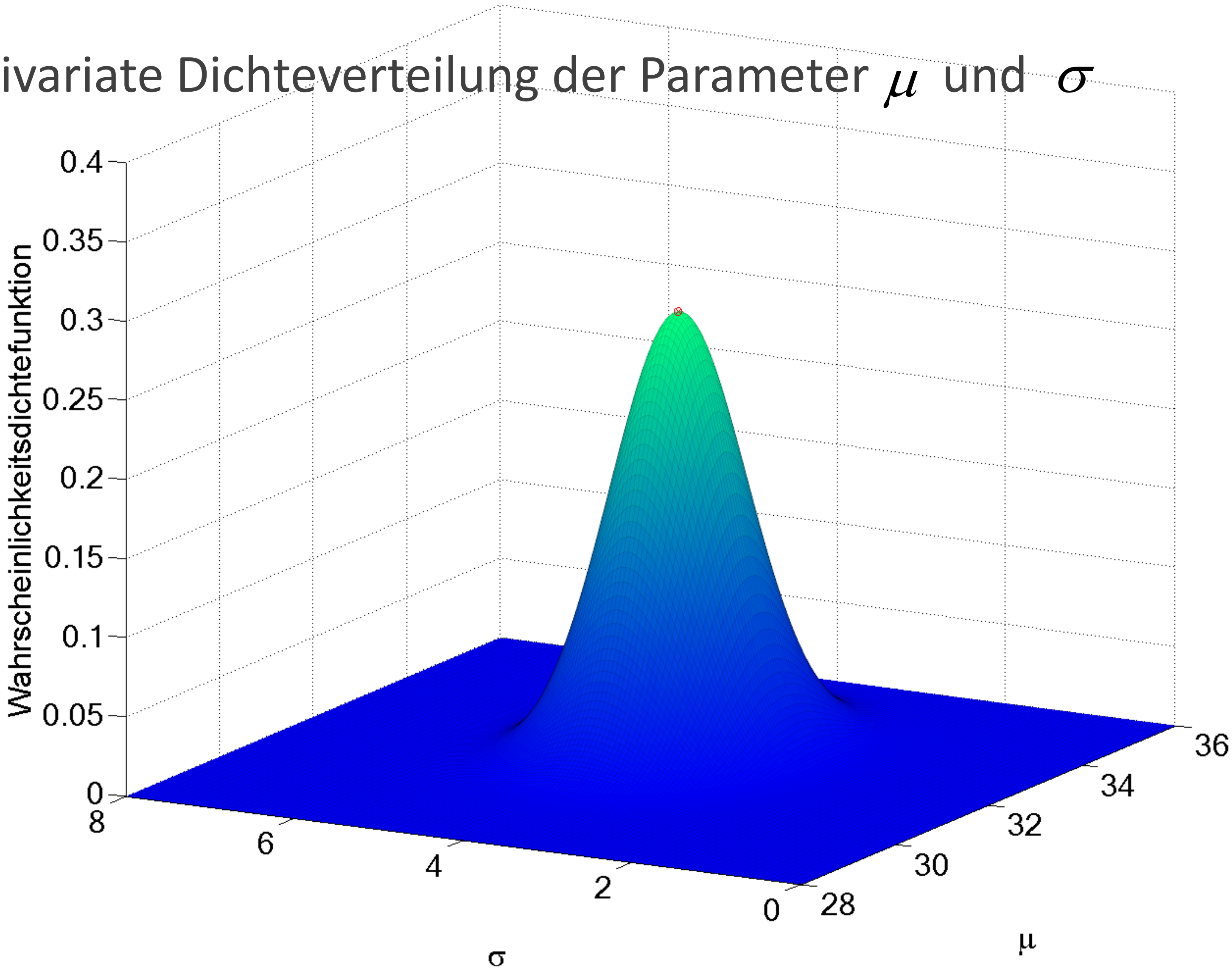
für Normalverteilung

$$\Downarrow$$

$$C_{\Theta\Theta} = H^{-1} = \begin{pmatrix} \text{Varianz von } \mu & \text{Kovarianz } \mu,\sigma \\ \text{Kovarianz } \sigma,\mu & \text{Varianz von } \sigma \end{pmatrix}$$

Aufgabe E.8

b) multivariate Dichteverteilung der Parameter μ und σ



Aufgabe E.8

c) Was passiert, wenn die Normalverteilung durch eine Exponentialverteilung ersetzt wird?

- Schätze die unbekannten Parameter der Exponentialfunktion mit Hilfe der MLM.
- Zeichne die kumulative Verteilungsfunktion zusammen mit der beobachteten Verteilungsfunktion der Stichprobenwerte.

Aufgabe E.8

Die Likelihood-Funktion ist:

Exponentialverteilung

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda \hat{x}_i}$$

Die korrespondierende Log-Likelihood-Funktion ist:

$$l = \ln L = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda \hat{x}_i}\right) = n \ln \lambda - \lambda \sum_{i=1}^{n} \hat{x}_i$$

Den MLM Schätzwert erhält man aus:

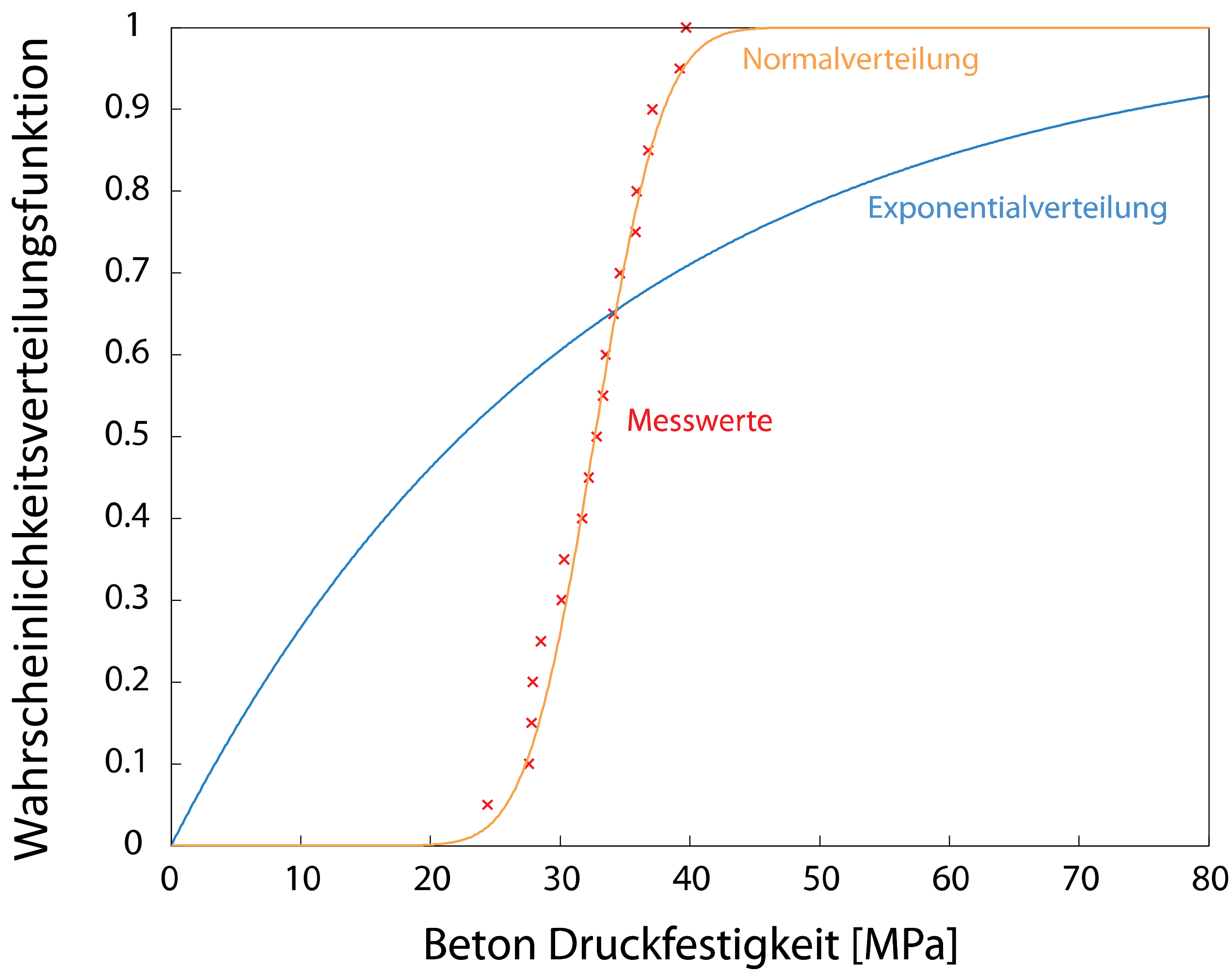
$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \hat{x}_i = 0$$

… er entspricht:

$$\lambda = \frac{n}{\sum_{i=1}^{n} \hat{x}_i} \longrightarrow \lambda = 0.031$$

Aufgabe E.8

c) Was meint ihr? Beschreibt die Exponentialfunktion (mit den MLM Schätzwerten) die Grundgesamtheit der Stichprobe gut?



Aufgabe E.11

Die Druckfestigkeit parallel zur Faser von 30 Holzproben wurde bestimmt. Die Ergebnisse sind in der Tabelle gegeben.

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

Aufgabe E.11

Die Druckfestigkeit parallel zur Faser von 30 Holzproben wurde bestimmt. Die Ergebnisse sind in der Tabelle gegeben.

a) Passe eine Exponential- und eine Weibullverteilung den Daten an. Bestimme dazu die Parameter dieser Verteilungen mit der Methode der Momente.

b) Zeichne die kumulativen Verteilungsfunktionen der Verteilungen und zeichne die kumulativen Verteilung der Stichprobe ein.

Aufgabe E.11

Die Druckfestigkeit parallel zur Faser von 30 Holzproben wurde bestimmt. Die Ergebnisse sind in der Tabelle gegeben.

a) Passe eine Exponential- und eine Weibullverteilung den Daten an. Bestimme dazu die Parameter dieser Verteilungen mit der Methode der Momente.

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

Aufgabe E.11

Das erste und zweite Moment der Beobachtungen sind:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = 26.41$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i^2 = 747.55$$

Die Exponentialverteilung hat die folgende Verteilung:

$$F_X(x) = 1 - \exp(-\lambda x), \quad x > 0$$

mit dem Mittelwert:

$$\mu = \frac{1}{\lambda} = m_1$$

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

Aufgabe E.11

Die Exponentialverteilung hat die folgende Verteilung:

$$F_X(x) = 1 - \exp(-\lambda x), \quad x > 0$$

mit dem Mittelwert: $\mu = \frac{1}{\lambda} = m_1$

So kann der Parameter λ geschätzt werden zu:

$$\lambda = \frac{1}{m_1} = 0.038$$

Die kumulative Verteilungsfunktion kann angegeben werden zu:

$$F_X(x) = 1 - \exp(-0.038x), \quad x > 0$$

Aufgabe E.11

Das erste und zweite Moment der Beobachtungen sind:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i = 26.41 \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n} \hat{x}_i^2 = 747.55 \qquad \mu = m_1 \qquad \sigma = \sqrt{m_2 - m_1^{\,2}}$$

Die (zweiparametrige) Weibullverteilung hat die folgende kumulative Verteilungsfunktion:

$$F_X(x) = 1 - \exp\left(-\left(\frac{x}{u}\right)^k\right), \quad x > 0$$

mit dem Mittelwert und der Standardabweichung:

$$\mu = u\Gamma\left(1 + \frac{1}{k}\right) = m_1$$

Siehe Skript Tabelle D.2

$$\sigma = u\sqrt{\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)} = \sqrt{m_2 - m_1^{\,2}}$$

Aufgabe E.11

Die unbekannten Parameter u und k können bestimmt werden:

1. Parameter k

Approximativ (ohne Beweis !) gilt:

(z.B. mit Solver aus Excel)

$$k \approx \left(\frac{\sigma}{\mu}\right)^{-1.09} = \left(\frac{\sqrt{m_2 - m_1^{\,2}}}{m_1}\right)^{-1.09} = 4.21$$

(Skript Tabelle D.2)

Aufgabe E.11

2. Parameter u:

$$\mu = u\Gamma\left(1+\frac{1}{k}\right)$$

$$u = \frac{\mu}{\Gamma(1+1/k)} = \frac{m_1}{\Gamma(1.24)}$$

$\Gamma(x)$	x	$\Gamma(x)$	x
.514	1.0	1.000	5.5
.863	1.1	0.951	5.6
.689	1.2	0.918	5.7
.811	1.3	0.897	5.8
.132	1.4	0.887	5.9
.591	1.5	0.886	6.0
.150	1.6	0.894	6.1

Linear interpolieren

Tab. T.5 Skript

Es werden die folgenden Parameter geschätzt:

$$k = 4.21 \qquad u = 29.05$$

Aufgabe E.11

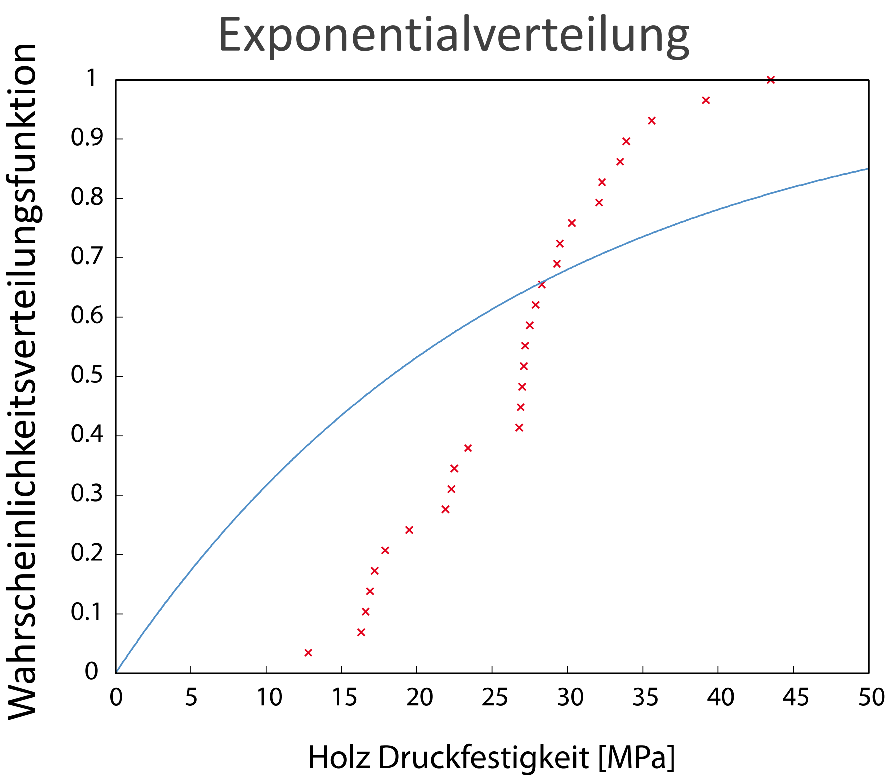
Die Druckfestigkeit parallel zur Faser von 30 Holzproben wurde bestimmt. Die Ergebnisse sind in der Tabelle gegeben.

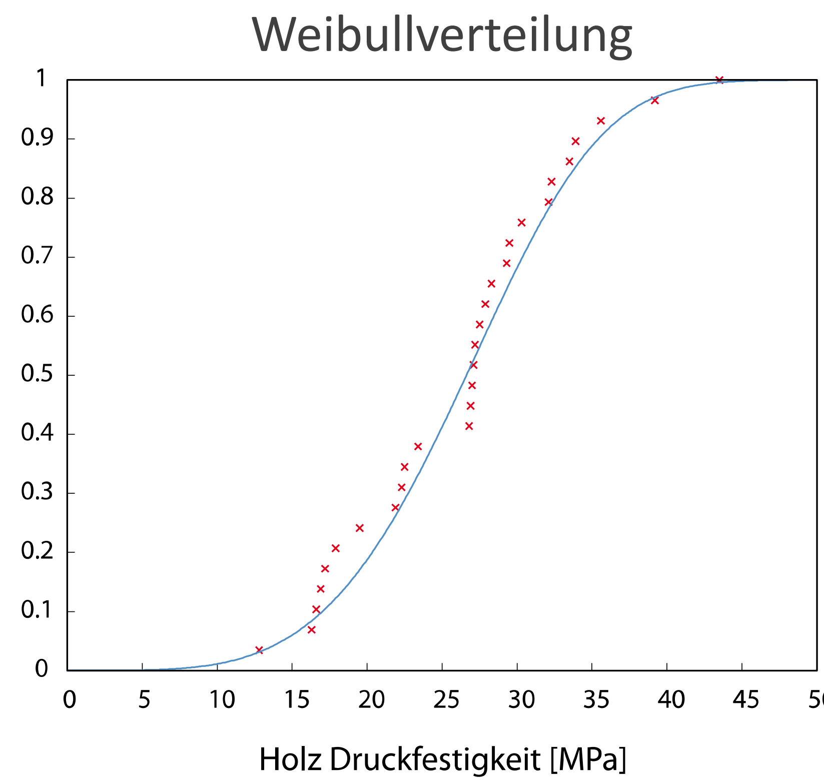
b) Zeichne die kumulative Verteilungsfunktion der beiden Verteilungen und zeichnen Sie jeweils die kumulative Verteilung der Stichprobe ein.

Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]	Nr.	Druck-festigkeit [MPa]
1	12.8	11	23.4	21	29.3
2	16.3	12	26.8	22	29.5
3	16.6	13	26.9	23	30.3
4	16.9	14	27	24	32.1
5	17.2	15	27.1	25	32.3
6	17.9	16	27.2	26	33.5
7	19.5	17	27.2	27	33.9
8	21.9	18	27.5	28	35.6
9	22.3	19	27.9	29	39.2
10	22.5	20	28.3	30	43.5

Aufgabe E.11

b) Zeichne die kumulative Verteilungsfunktion der beiden Verteilungen und zeichne jeweils die kumulative Verteilung der Stichprobe ein.





Hausübung E.6

Zur Dimensionierung eines Parkhauses wurden die Ankunftszeiten von Fahrzeugen aufgenommen. Die Zeitdifferenzen zwischen ankommenden Fahrzeugen sind in der Tabelle gegeben.

a) Zeichne das Wahrscheinlichkeitspapier für die Exponentialverteilung und trage die Werte der Zeitdifferenzen ein. Beurteile, ob die Intervalle einer Exponentialverteilung folgen.

b) Bestimme den Erwartungswert der Zeitdifferenz grafisch aus der in a) erzeugten Grafik unter der Annahme, dass die Daten exponentialverteilt sind. Berechne den Erwartungswert und vergleiche die beiden Werte.

i	Zeitdifferenz x_i^o (Sekunden)
1	1.52
2	6.84
3	9.12
4	10.64
5	15.20
6	21.28
7	30.40
8	30.40
9	34.20
10	60.80
11	78.28
12	95.76

Verwende für die beobachtete kumulative Wahrscheinlichkeitsverteilungsfunktion: $F_X(\hat{x}_i^o) = \dfrac{i}{N+1}$

Hausübung E.6

b) Bestimme den Erwartungswert der Zeitdifferenz grafisch aus der in a) erzeugten Grafik unter der Annahme, dass die Daten exponentialverteilt sind. Berechne den Erwartungswert und vergleiche die beiden Werte.

1) Bestimme λ über die Steigung der Geraden

2) Beziehung zwischen μ und λ für Exponentialverteilung:

$$\mu = \frac{1}{\lambda}$$

3) Erwartungswert der Zeitdifferenz:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_i$$

i	Zeitdifferenz $\overset{o}{x}_i$ (Sekunden)
1	1.52
2	6.84
3	9.12
4	10.64
5	15.20
6	21.28
7	30.40
8	30.40
9	34.20
10	60.80
11	78.28
12	95.76