

Benjamin Klein

**Numerische Analyse
von gemischten Ausfallverteilungen
in der Zuverlässigkeitstechnik**

D 93
ISBN 978-3-936100-49-5

Institut für Maschinenelemente

Antriebstechnik • CAD • Dichtungen • Zuverlässigkeit

Universität Stuttgart
Pfaffenwaldring 9
70569 Stuttgart
Tel. (0711) 685 – 66170

Prof. Dr.-Ing. B. Bertsche, Ordinarius und Direktor

Numerische Analyse von gemischten Ausfallverteilungen in der Zuverlässigkeitstechnik

Von der Fakultät
Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von

Dipl.-Ing. Benjamin Klein
geboren in Bietigheim-Bissingen

Hauptberichter: Univ.-Prof. Dr.-Ing. B. Bertsche
Mitberichter: Univ.-Prof. Dr. rer. nat. U. Jensen

Tag der mündlichen Prüfung: 26. Juli 2013

Institut für Maschinenelemente der Universität Stuttgart
2013

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als akademischer Mitarbeiter am Institut für Maschinenelemente (IMA) der Universität Stuttgart.

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr.-Ing. Bernd Bertsche, Leiter des Instituts für Maschinenelemente, für die Ermöglichung und Förderung meiner Arbeit und das mir stets entgegengebrachte Vertrauen.

Herrn Prof. Dr. rer. nat. Uwe Jensen, Institut für Angewandte Mathematik und Statistik der Universität Hohenheim, danke ich für die Übernahme des Mitberichts, die kritische Durchsicht und die konstruktiven Hinweise.

Ebenso gilt mein Dank Herrn Prof. Dr.-Ing. habil. Werner Haas, Leiter des Bereichs Dichtungstechnik am IMA für die Betreuung während meiner Anfangszeit am IMA bei der Bearbeitung des DFG-Projekts „Zuverlässigkeit von Radial-Wellendichtungen“.

Mein herzlicher Dank gilt allen aktuellen und ehemaligen Mitarbeitern des Instituts für das sehr freundschaftliche Arbeitsklima und die stete Hilfsbereitschaft über die Bereiche und die eigentliche Arbeitszeit hinaus. Zudem gilt mein Dank auch den wissenschaftlichen Hilfskräften, den Studien- und Diplomarbeitern die mit zum Gelingen dieser Arbeit beigetragen haben.

Metterzimmern, im August 2013

Benjamin Klein

Inhalt

Bezeichnungen und Formelzeichen	v
Abkürzungen	x
Abstract	xi
1 Einleitung.....	1
1.1 Problemstellung	2
1.2 Ziele der Arbeit	3
1.3 Aufbau der Arbeit	3
2 Stand der Forschung und Technik.....	5
2.1 Grundlagen, Begriffe und Definitionen in der Zuverlässigkeitstechnik.....	5
2.2 Verteilungen und Systemtheorie.....	8
2.2.1 Weibullverteilung.....	9
2.2.2 Gemischte Verteilungen.....	10
2.2.3 Abgrenzung von gemischten Verteilungen.....	14
2.3 Parameterschätz- und Optimierungsverfahren.....	16
2.3.1 Parameterschätzung von Einzelverteilungen	16
2.3.2 Graphische Methoden zur Parameterschätzung von Mischverteilungen.....	18
2.3.3 Analytische Methoden zur Parameterschätzung von Mischverteilungen.....	21
2.3.4 Numerische Optimierungsmethoden	23
2.4 Monte-Carlo Methode.....	26
3 Grundlagen numerischer Optimierungsmethoden	28
3.1 Grundlagen genetischer Algorithmen	28
3.1.1 Varianten der evolutionären Algorithmen	28
3.1.2 Anwendung evolutionärer Algorithmen in der Zuverlässigkeitstechnik.....	29
3.1.3 Begriffe der genetischen Algorithmen.....	30
3.1.4 Ablauf des genetischen Algorithmus	31
3.1.5 Fitnessfunktion.....	33
3.1.6 Kodierung.....	33
3.1.7 Genetische Operatoren.....	34
3.1.8 Selektion.....	36
3.1.9 Konvergenzverhalten des genetischen Algorithmus.....	38
3.2 Grundlagen künstlicher neuronaler Netze	39
3.2.1 Neuronale Netze in der Zuverlässigkeitstechnik	40

3.2.2	Aufbau eines neuronalen Netzwerkes	41
3.2.3	Das Neuron.....	42
3.2.4	Netztraining	43
3.2.5	Lernverfahren	44
3.2.6	Vorgehensweise bei der Parameterschätzung von realen Daten.....	46
3.3	Vergleich der Anwendung von genetischen Algorithmen und neuronalen Netzen.....	47
4	Parameterschätzung mit genetischen Algorithmen.....	49
4.1	Mischverteilungen zur Evaluation der Parameter des genetischen Algorithmus.....	51
4.2	Definition der Kodierung	53
4.3	Definition der Fitnessfunktion	58
4.3.1	Fitness in Ordinatensrichtung	58
4.3.2	Fitness in Abszissenrichtung.....	61
4.3.3	Fitness mit der Methode der minimalen Abstandsquadrate.....	63
4.3.4	Likelihoodfunktion als Fitnessfunktion	65
4.3.5	Vergleich der Varianten zur Fitnessbestimmung.....	66
4.3.6	Gewichtung der Datenpunkte.....	67
4.4	Optimierung der Parameter des genetischen Algorithmus	73
4.4.1	Rekombination	73
4.4.2	Mutation	76
4.4.3	Selektion.....	78
4.4.4	Einfluss der Generationenanzahl, Populations- und Suchraumgröße.....	82
4.5	Zusammenfassung.....	85
5	Gemischte Verteilungen zur Evaluation.....	87
5.1	Mischverteilungen.....	88
5.1.1	Mischverteilungen nach Jiang und Kececioglu.....	88
5.1.2	Abgeleitete Mischverteilungen mit mehreren Unterverteilungen.....	90
5.1.3	Mischverteilungstypen von Jiang und Kececioglu erweitert um t_0	93
5.2	Boolesche Systeme.....	95
5.2.1	Separierbarkeit.....	95
5.2.2	Datensätze der booleschen Seriensysteme	100
5.2.3	Datensätze der booleschen Parallelsysteme	101
5.3	Zensierte Datensätze	103
6	Parameterschätzung von gemischten Verteilungen.....	104
6.1	Methode zum Vergleich der geschätzten Verteilungen	104
6.2	Parameterschätzung von Einfachverteilungen	106
6.3	Parameterschätzung von Mischverteilungen.....	107

6.3.1	Parameterschätzung der Mischverteilungen nach Jiang und Kececioglu	107
6.3.2	Parameterschätzung von Mischverteilungen mit mehreren Unterverteilungen.....	110
6.3.3	Parameterschätzung der abgeleiteten Mischverteilungen mit t_0	115
6.4	Parameterschätzung von Systemen.....	117
6.4.1	Parameterschätzung von Seriensystemen	117
6.4.2	Parameterschätzung von Parallelsystemen	120
7	Schätzung der Unterverteilungsanzahl und Bewertung der Anwendung.....	122
7.1	Schätzung der Unterverteilungsanzahl	122
7.1.1	Schätzung der Unterverteilungen in Mischverteilungen.....	123
7.1.2	Schätzung der Unterverteilungen in Seriensystemen	125
7.1.3	Möglichkeiten zur Verbesserung der Schätzung der Unterverteilungsanzahl	127
7.2	Bewertung der Anwendung	129
8	Zusammenfassung und Ausblick	131
9	Literatur	133
10	Anhang.....	143
10.1	Verwendete Parameter des genetischen Algorithmus	143
10.2	Mittels Monte-Carlo Methode erzeugte Datensätze	148

Bezeichnungen und Formelzeichen

a	Trainingsdatensatz
a_i	Abstand zwischen benachbarten Datenpunkten
b	Formparameter der Weibullverteilung
B_{10}	Lebensdauer bei einer Ausfallwahrscheinlichkeit von 10 %
b_a	Formparameter Verteilung a
$b_{\text{Abweichung}}$	abweichender Formparameter der geschätzten Verteilung
b_b	Formparameter Verteilung b
b_g	geschätzter Formparameter
b_i	Formparameter der i -ten Unterverteilung
C	Anzahl der Kanten im künstlichen neuronalen Netz
c_{ij}	j -te Kante in i -ter Schicht
D_F	Abstand in Ordinatenrichtung zur Berechnung der kleinsten Abstandsquadrate
D_h	Abstand in Abszissenrichtung
d_i	Anzahl der Ausfälle im Zeitintervall i
D_i	Abstand des Punktes i von der geschätzten Kurve
D_{krit}	kritischer Wert des Kolmogorov-Smirnov-Anpassungstest
D_{max}	maximaler Ordinatenabstand
D_{median}	Medianwert des Ordinatenabstands
D_q	Abstand berechnet mit minimalen Abstandsquadraten
D_S	Summe der minimalen Abstandsquadrate
D_t	Abstand in Abszissenrichtung zur Berechnung der kleinsten Abstandsquadrate
D_v	Abstand in Ordinatenrichtung
E_{max}	Erwartungswert des besten Individuums
E_{min}	Erwartungswert des schlechtesten Individuums
F'	transformierte Ausfallwahrscheinlichkeit für Weibullnetz
$F(t)$	Ausfallwahrscheinlichkeit
$f(t)$	Dichtefunktion
F_0	theoretische Verteilungsfunktion
$F_{1,2}$	Schnittpunkt zwischen erster und zweiter Unterverteilung
$F_{2,3}$	Schnittpunkt zwischen zweiter und dritter Unterverteilung
F_B	empirische Verteilung der Stichprobe

F_E	theoretische Verteilung der Stichprobe
F_i	Ausfallwahrscheinlichkeit der i -ten Ranggröße
F_i	Ausfallverteilung der i -ten Unterverteilung
$F_i(t)$	Ausfallwahrscheinlichkeit der i -ten Komponenten
$f_i(t)$	Dichtefunktion der i -ten Unterverteilung
f_M	Dichtefunktion der Mischverteilung
$F_M(t)$	Ausfallverteilung der Mischverteilung
F_M^{-1}	Inverse der Mischverteilung
$F_{median}(t_i)$	Medianrang zum Zeitpunkt t_i
F_n	empirische Verteilungsfunktion
F_S	Systemlebensdauer
F_S^{-1}	Inverse der Systemfunktion
f_{tb}	Verhältniszahl zur Bestimmung der ausfallfreien Zeit
G	Anzahl der Generationen
g	Anzahl der vor dem Ausfall liegenden Teile
G_1	erste Tangente an Datenpunkte
G_2	zweite Tangente an Datenpunkte
H_0	Nullhypothese
H_1	Horizontale bei Jensen-Petersen-Methode
H_2	Horizontale bei Jensen-Petersen-Methode
H_A	Alternativhypothese
i	Rangzahl, Zählvariable
I	Individuum
I_j	j -tes Individuum
I_k	k -tes Individuum
$j(t_i)$	hypothetische Rangzahl
k	Mindestanzahl an Datenpunkten je Verteilung
L	Likelihoodfunktion
L_i	Kodelänge des Parameters i
L_K	Kodierungslänge
L_T	Lösungsraum Parameter T
L_x	Kodelänge des Parameters x
m	Geradensteigung
M	Menge der Lösungsvarianten
m_T	Genauigkeit Parameter T
m_x	Genauigkeit des Parameters x

n	Stichprobenumfang
N	Anzahl Systemkomponenten, Anzahl Unterverteilungen
N_C	Anzahl der Crossover-Punkte
$N(t_j)$	Zuwachs
net	Propagierungsfunktion
n_i	Anzahl der zum Zeitpunkt t_i noch nicht ausgefallenen Teile
n_{ji}	j -ter Knoten/Neuron in i -ter Schicht
N_K	Anzahl der Knoten/Neuronen im künstlichen neuronalen Netz
n_P	Anzahl der Parameter je Verteilung
n_{va}	Anzahl an hinreichenden Parameterschätzungen für Verteilung a
n_{vb}	Anzahl an hinreichenden Parameterschätzungen für Verteilung b
o_i	Ausgangswert Neuron i
\bar{o}_j	Sollausgabe des Neurons j
P	Wahrscheinlichkeit
p_{cross}	Rekombinationswahrscheinlichkeit für Uniform-Crossover
p_i	Gewichtungsfaktor der i -ten Unterverteilung
$p_{i, Norm}$	Normierter Gewichtungsfaktor
p_{mut}	Mutationswahrscheinlichkeit
p_S	Selektionswahrscheinlichkeit
Q	Energiefunktion
$r(I)$	Rangzahl des Individuums I
$R(t)$	Zuverlässigkeit, Überlebenswahrscheinlichkeit
r_i	minimale Entfernung Punkt und Funktionsgraph
S	Schwellwert der Aktivierungsfunktion
t	Zeit, Lebensdauermerkmal
T	charakteristische Lebensdauer der Weibullverteilung
t'	transformierte Ausfallzeit für Weibullnetz
t_0	ausfallfreie Zeit der Weibullverteilung
t_{0i}	ausfallfreie Zeit der i -ten Unterverteilung
t_{0o}	obere Intervallgrenze der ausfallfreien Zeit
t_{0u}	untere Intervallgrenze der ausfallfreien Zeit
T_a	charakteristische Lebensdauer Verteilung a
$T_{Abweichung}$	abweichende charakteristische Lebensdauer der geschätzten Verteilung
T_b	charakteristische Lebensdauer Verteilung b
t_i	Ausfallzeit der i -ten Ranggröße
T_i	charakteristische Lebensdauer der i -ten Unterverteilung

T_o	obere Intervallgrenze der charakteristischen Lebensdauer
T_u	untere Intervallgrenze der charakteristischen Lebensdauer
u	Anzahl der Parameter
u_I	Anzahl der Individuen
\underline{w}	Gewichtungsvektor der Fitness
w_i	Gewichtung des i -ten Datenpunktes
w_{ij}	Gewichtung der Kante i, j
$x(t_i)$	Abszissenwert zur Abstandsbestimmung
$X_{(y)}$	Wert der binären Kodierung an der Stelle y
$X_{G(y)}$	Wert der Gray- Kodierung an der Stelle y
x_{bin}	Wert von x im Binärkode
x_{bin+1}	Wert x im Binärkode um eine Stelle verschoben
x_{dec}	Dezimalwert des Parameters x
x_{Gray}	Wert von x im Gray-Kode
x_i	Abszissenwert von Punkt i
x_o	obere Intervallgrenze des Parameters x
$x_{o, i}$	obere Intervallgrenze von Parameter x aus Unterverteilung i
x_P	Abszissenwert des Punktes P
x_u	untere Intervallgrenze des Parameters x
$x_{u, i}$	untere Intervallgrenze von Parameter x aus Unterverteilung i
$y(F_i)$	Ordinatenwert zur Abstandsbestimmung
y_a	Achsenabschnitt
y_i	Ordinatenwert von Punkt i
y_P	Ordinatenwert des Punktes P
Z	Ziel-/ Fitnessfunktion des genetischen Algorithmus
$Z(I_j)$	Fitness j -tes Individuum
$Z_{h, ges}$	Fitnesswert in Abszissenrichtung
$Z_{q, ges}$	Fitnesswert aus minimalen Abstandsquadraten
$Z_{v, ges}$	Fitnesswert in Ordinatenrichtung
α	Parameter der Sigmoidfunktion
δ	Differenz zwischen Zielausgabe und aktueller Ausgabe
$\Delta_a w_{ij}$	Gewichtungsänderung durch Trainingsdatensatz a
ζ	Populationsgröße
η	Lernrate

κ	gleichverteilte Zufallsgröße
λ	Lagrange-Multiplikator
$\lambda(t)$	Ausfallrate
μ	Parameter des Momentumterms
ν	Anzahl der Parameter der Verteilung
ρ	Summe der Abstandquadrate
ρ_L	theoretische Lösungsdichte
σ	Verhältnis der Parameterschätzungen
τ	Lebensdauer
Φ	Aktivierungsfunktion
φ_{opt}	optimale Parameter der Verteilung
ψ	Vektor der Verteilungsparameter

Abkürzungen

GA	genetischer Algorithmus
gew.	gewichtet
K	Komponente boolesches System
KNN	künstliches neuronales Netz
K-S-Test	Kolmogorov-Smirnov-Anpassungstest
Max	Maximalwert
MC	Monte-Carlo Methode
Med	Medianwert
Min	Minimalwert
Mischvert.	Mischverteilung
MLE	Maximum-Likelihood Estimation/ Maximum-Likelihood Methode
MLE/EM	Maximum-Likelihood Methode mit Expectation-Maximization Algorithmus
Moment.	Momentenmethode
MW	arithmetischer Mittelwert
Regr.	Regressionsmethode
Untervert.	Unterverteilung
UV	Unterverteilung
W++	Software Weibull++ [®] von Reliasoft [®]

Abstract

Numerical analysis of mixture distributions in reliability

The verification of reliability targets is expensive in terms of time and money. Two options for proving reliability targets exist, bench testing or analysis of field data. Both possibilities have their advantages and disadvantages. Tests offer the advantage of knowing the testing parameters precisely, usually at the expense of small sample sizes and synthetic testing conditions. Compared to bench tests, field data has the advantage of realistic loads on the product and large sample sizes, with the disadvantage of uncertainty in actual lifetime and stress information. Despite these disadvantages, field data analysis has a great potential for determining real lifetime and failure mechanisms of products.

Field datasets often consist of a combination of different product types or failure mechanisms. These combinations can occur due to mixed distributions or failures of various components of a system. In this work, a new method to analyze these mixtures of distributions is presented.

Firstly the known graphical and analytical methods for analyzing mixed distributions are illustrated. The limitations of the methods mentioned before are discussed concerning the analysis of mixed distributions and Boolean systems. Due to the limitations of analytical methods, numerical optimization methods and their ability in estimating distribution parameters are discussed.

The most promising optimization methods were chosen and analyzed with regard to the present problem. Genetical algorithms and artificial neural networks were examined in detail for applicability to the problem. The function, structure and application of both optimization methods were described and discussed. Both methods were compared relating their applicability to solve the problem of parameter determination of mixture distributions. As a result of the comparison, the genetic algorithms were determined to be the better option, in terms of both versatility and easy implementation.

The genetic algorithm to determine the parameters of mixture distributions was implemented in Matlab. Using exemplary representative mixed distributions, the different parameters of the algorithm were analyzed and optimized. Therefore the encoding of the individuals, the applicable fitness function and the optimal parameters of the algorithm were defined in the next step. Based on this analysis, a set of specifications for the parameter estimation of mixture distributions could be recommended.

The functionality and expected quality of the results of the genetic algorithm was evaluated using several representative datasets. Genetic algorithms, such as a stochastic method, have the ability to find a global optimum for the given problem. Therefore large population sizes and high numbers of generations are necessary. On the basis of the presented datasets, the results that can be expected with a limited population size and number of generations were analyzed. To achieve this, the representative datasets should include as many types of mixed distributions and Boolean systems as possible. Based on this, the results can, in general, be considered applicable.

After defining the datasets, the distribution parameters were determined using genetic algorithms and the results were discussed. The achieved results for the mixed distributions without failure-free time of the genetic algorithm are in the same range as the results of commercial software products. As expected, increasing numbers of sub-distributions lead to a decreasing accuracy of the results. Parameter determinations of mixed distributions with failure free time and Boolean systems show good results. The final section deals with the determination of the number of sub-distributions. Sub-distributions were considered as an additional parameter to be determined. The accuracy of the results was discussed and different points for improvement actions were established.

With the genetic algorithm a method was implemented, that is applicable to determine distribution parameters of all types of mixture distributions. The accuracy of the results is close to the ones of the analytical method of comparison the MLE/EM-Algorithm. Compared to the analytical method, the universal applicability of only one procedure to all mixture distributions compensates for the disadvantage of slightly worse results.

The genetic algorithm presented in this work shows good results for the evaluation of the presented datasets. A general verification for the applicability to other types of distributions, such as the normal or lognormal distribution, is pending. In addition, further extensions of the genetic algorithm are known, as for example memetic algorithms which have not yet been studied. This work showed that numerical methods have the principal ability for the parameter estimation of mixture distributions. Based on this, further numerical methods, such as artificial neural networks, could be considered for their applicability to this problem.

1 Einleitung

Eine der wichtigsten Anforderungen an ein Produkt ist seine Zuverlässigkeit. Obwohl meist nicht als explizite Anforderung von Kunden angegeben, ist sie doch als implizite Anforderung vorhanden. Dabei wird eine der Produktkategorie, beispielsweise Automobile, entsprechende lange Lebensdauer ohne Einbußen der Funktionalität erwartet. Um diese impliziten Anforderungen umzusetzen, werden schon in frühen Phasen der Produktentstehung Zuverlässigkeitszielgrößen festgelegt [1].

Aussagen über die Zuverlässigkeit können einerseits durch Versuche nachgewiesen werden, andererseits aus der Analyse von Felddaten gewonnen werden. In beiden Fällen ist für eine belastbare Aussage ein großer Aufwand erforderlich. Entweder müssen viele Teile im Versuch bis zum Ausfall getestet werden, oder es müssen ausgefallene Teile aus dem Feld befundet und analysiert werden. Der Vorteil der Felddaten gegenüber denen aus Versuchen liegt in der meist höheren Anzahl an Ausfällen. Zudem wirken im Feld jene Belastungen auf ein Produkt, die im normalen Betrieb auftreten können. Versuchsergebnisse dagegen sind direkt abhängig von der gewählten Belastung bzw. dem Belastungskollektiv und können dementsprechend nur einen Fall nahe der eigentlichen Belastung im Betrieb abbilden. Die Analyse von Felddaten bietet daher ein großes Potential, um die Ausfallmechanismen und Lebensdauerverteilung eines Produktes unter realistischen Bedingungen zu bestimmen [2]. Die so gewonnenen Erkenntnisse können in verschiedenen Phasen des Produktlebenszyklus des aktuellen Produkts oder eines Nachfolgeprodukts genutzt werden; einerseits um Schwachstellen zu beseitigen, andererseits zur Abschätzung der Garantiekosten [3].

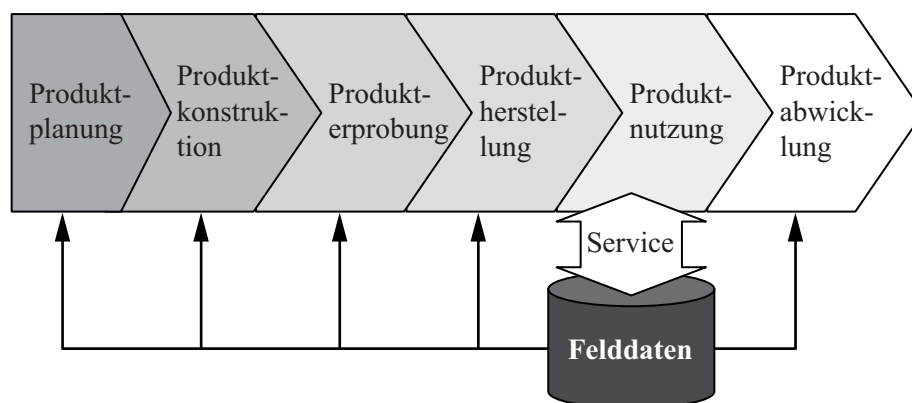


Bild 1.1: Nutzung von Felddaten im Produktlebenszyklus [2]

1.1 Problemstellung

Die Zuverlässigkeit eines Produkts wird zu einem großen Teil von seiner Komplexität bestimmt [1]. Die zunehmende Integration von Funktionen in ein Produkt führt zu einer Erhöhung seiner Komplexität [4]. Eine Steigerung der Komplexität bewirkt zusätzlich meist eine Zunahme an Ausfallmechanismen, da jede Komponente, die in einem Produkt verbaut ist, eigene spezifische Ausfallmechanismen besitzt. Dadurch steigt mit der Anzahl der Komponenten auch die Anzahl der Ausfallmechanismen.

Je komplexer ein Produkt ist, umso schwieriger wird demnach auch die Analyse von Ausfalldaten, da eine Zuordnung des jeweiligen Ausfalls zur zugehörigen Komponente bzw. zum zugehörigen Ausfallmechanismus erfolgen muss. Insbesondere bei der Analyse von Felddaten ist eine Differenzierung der einzelnen Ausfälle nach ihren Ausfallmechanismen meist nicht möglich. Die erhobenen Felddaten sind nicht ausreichend dokumentiert oder nicht aussagekräftig genug, um eine klare Differenzierung zu erreichen. Dadurch können die Ausfallmechanismen nicht getrennt voneinander ausgewertet werden. Stattdessen bestehen die vorliegenden Felddaten aus einem Gemisch an Ausfällen mit unterschiedlichen Ausfallverteilungen. Hierbei können verschiedene Produkttypen aus verschiedenen Baureihen, beispielsweise aus Getrieben eines Baukastensystems, in einem Ausfalldatensatz vorhanden sein. Dies führt zu einer Mischverteilung. Demgegenüber können auch verschiedene Ausfallmechanismen in einem Produkt auftreten, die unsortiert im Ausfalldatensatz vorhanden sind. Beispiele hierfür sind Getriebe, die als Seriensysteme mit unterschiedlichen Komponenten mit unterschiedlichen Ausfallverteilungen behandelt werden können.

Eine Abbildung dieser Daten mit einer einzelnen Verteilung führt oftmals nicht zu einer befriedigenden Übereinstimmung der Verteilungsfunktion mit dem Ausfalldatensatz. Stattdessen muss, um eine hinreichende Übereinstimmung zu erreichen, eine Mischverteilungsanalyse durchgeführt werden. Mit Hilfe der Mischverteilungsanalyse können die Parameter von verschiedenen, den Ausfalldaten zugrunde liegenden Verteilungen geschätzt werden. Die bekannten graphischen und analytischen Methoden sind oftmals Restriktionen unterworfen, die eine Anwendung erschweren. Die graphischen Methoden nach Kao und Cran [5], [6] und Jensen und Petersen [7] können beispielsweise nur auf Mischverteilungen mit zwei Unterverteilungen angewandt werden. Die analytischen Methoden können bei Mischverteilungen, die Unstetigkeitsstellen, beispielsweise durch eine ausfallfreie Zeit, beinhalten, nicht angewandt werden.

Ein neuer Ansatz ist es, numerische Methoden zur Verteilungsparameterschätzung in der Zuverlässigkeitstechnik zu verwenden. Häufig werden diese Methoden zur Optimierung von komplexen Problemen angewandt. Das populärste Beispiel ist das Problem des Handlungsreisenden, siehe Bild 1.2 [8]. Hierbei muss der kürzeste Rundkurs

zwischen verschiedenen vorgegebenen Orten, in Bild 1.2 als Punkte dargestellt, ermittelt werden. Die Schwierigkeit des Problems steigt mit zunehmender Anzahl an Orten, die zu dem Rundkurs verbunden werden sollen, überexponentiell an.

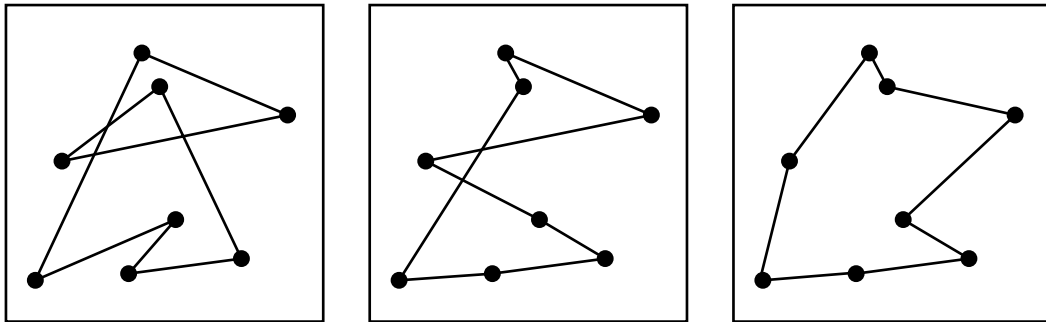


Bild 1.2: Problem des Handlungsreisenden

Durch die mittlerweile hohe verfügbare Rechenleistung moderner Arbeitsplatzcomputer sind diese rechenintensiven numerischen Methoden alltagstauglich geworden. Komplexe Optimierungsaufgaben lassen sich somit mit einem geringen Aufwand lösen.

1.2 Ziele der Arbeit

Die vorliegende Arbeit soll die grundsätzliche Eignung von numerischen Methoden zur Schätzung von Verteilungsparametern in gemischten Verteilungen aufzeigen. Als gemischte Verteilungen werden in dieser Arbeit einerseits Mischverteilungen aus gewichteten Unterverteilungen und andererseits Systemausfallverteilungen von Systemen nach der booleschen Systemtheorie angesehen. Das Hauptaugenmerk der verwendeten numerischen Methoden liegt auf evolutionären Algorithmen und hier auf deren Unterart, den genetischen Algorithmen. Diese Methode kann den Heuristiken oder stochastischen Optimierungsverfahren zugeordnet werden. Als weitere Methode werden künstliche neuronale Netze und ihre Anwendung auf diese Problemstellung diskutiert und das Verfahren beispielhaft beschrieben.

1.3 Aufbau der Arbeit

Im Kapitel **Stand der Forschung und Technik** werden die gebräuchlichen Begrifflichkeiten und Definitionen der Zuverlässigkeitstechnik erläutert und in die verschiedenen verwendeten Arten von Verteilungen, Einfachverteilungen und gemischte Verteilungen, eingeführt. Zudem werden verschiedene Verfahren zur Parameterschätzung und Optimierung vorgestellt und nach ihrer Methodik in graphische, analytische und numerische Verfahren eingeteilt. Abschließendes Thema des Kapitels ist die Monte-Carlo Methode, die zur näherungsweise Lösung von analytischen Problemen verwendet werden kann.

Im anschließenden Kapitel **Grundlagen der numerischen Optimierungsverfahren** werden die zwei vielversprechendsten numerischen Methoden, die evolutionären Algorithmen und die künstlichen neuronalen Netze, beschrieben. Dabei werden sowohl der Aufbau als auch der Ablauf zur Parameterschätzung aufgezeigt und die zentralen Einstellgrößen der jeweiligen Methode erläutert. Abschließend werden beide Methoden miteinander verglichen und die Eignung zur Parameterschätzung von gemischten Verteilungen diskutiert.

Der Hauptteil der Arbeit findet sich in Kapitel **Parameterschätzung mit genetischen Algorithmen**. Hier wird der Ablauf des genetischen Algorithmus für die Parameterschätzung von gemischten Verteilungen beschrieben. Anschließend werden die verschiedenen zentralen Einstellgrößen des Verfahrens aufgezeigt und gezielt auf die Anwendung zur Schätzung von Verteilungsparametern von gemischten Verteilungen hin optimiert.

Das Kapitel **gemischte Verteilungen zur Evaluation** beschreibt die zur Evaluation der Funktionalität des genetischen Algorithmus eingesetzten Datensätze. Die Mischverteilungsdatensätze basieren auf den verschiedenen von Jiang und Kececioğlu [41] aufgezeigten Typen von Mischverteilungen. Des Weiteren werden die Überlegungen zur Separierbarkeit von booleschen Systemen und die daraus abgeleiteten Datensätze vorgestellt. Die Datensätze sind für die Evaluation der Schätzungsgüte der eingesetzten Verfahren maßgeblich, weshalb auf die Auswahl der verwendeten Datensätze ausführlich eingegangen wird.

Im nachfolgenden Kapitel **Parameterschätzung von gemischten Verteilungen** werden die Parameter der im vorangegangenen Kapitel beschriebenen repräsentativen gemischten Verteilungen mittels des genetischen Algorithmus geschätzt. Zum Vergleich der Ergebnislänge kommen, soweit möglich, verschiedene analytische Methoden zum Einsatz. Die Schätzungen werden erläutert und Abweichungen bei der Parameterschätzung diskutiert.

Das Kapitel **Schätzung der Unterverteilungsanzahl und Bewertung der Anwendung** zeigt die Möglichkeiten der Schätzung der Unterverteilungsanzahl von Mischverteilungen und Seriensystemen auf. Anhand von beispielhaften Schätzungen werden die Problematiken bei Mischverteilungen und Systemen betrachtet und diskutiert. Zuletzt wird die Anwendbarkeit des genetischen Algorithmus bei der Schätzung von gemischten Verteilungen beschrieben und diskutiert. Dazu werden die zentralen Punkte der Parameterschätzung aufgezeigt.

2 Stand der Forschung und Technik

Die Schätzung von Verteilungsparametern ist eine häufige Problemstellung in der Zuverlässigkeitstechnik. Verteilungsparameter führen zu einer besseren Vergleichbarkeit von verschiedenen Datensätzen, als es einzelne Datenpunkte an sich vermögen. In diesem Abschnitt werden die Vorgehensweisen zur Schätzung von Parametern dargestellt.

Im allgemeinen Teil werden zuerst die Grundlagen und Definitionen der in der Zuverlässigkeitstechnik verwendeten Begriffe erläutert. Nachfolgend werden die für die Arbeit verwendete Weibullverteilung und die verschiedenen Arten der gemischten Verteilungen vorgestellt und voneinander abgegrenzt. Den Hauptteil bilden die verschiedenen Methoden zur Schätzung von Parametern in gemischten Verteilungen. Diese lassen sich in drei Hauptgruppen unterteilen:

- graphische Methoden
- analytische Methoden
- numerische Methoden

Die Beschreibung der Monte-Carlo Methode, die für die Erzeugung von Datensätzen erforderlich ist, bildet den letzten Teil.

2.1 Grundlagen, Begriffe und Definitionen in der Zuverlässigkeitstechnik

In diesem Kapitel werden die grundlegenden Begrifflichkeiten und Definitionen in der Zuverlässigkeitstechnik, basierend auf den Standardwerken [1], [10] - [22], vorgestellt.

Ausfall

Der Ausfall ist in der Zuverlässigkeitstechnik nach [23] als „*Beendigung der Fähigkeit einer Einheit, eine geforderte Funktion zu erfüllen*“ definiert. Er ist die Grundlage für die weitere Untersuchung der Ausfallwahrscheinlichkeit.

Ausfallwahrscheinlichkeit

Die Ausfallwahrscheinlichkeit $F(t)$ ist definiert als eine monoton steigende Verteilungsfunktion. Dabei ist $F(t)$ die Wahrscheinlichkeit für einen Ausfall im Intervall $[0, t]$. Es gilt:

$$F(t) = P(\tau \leq t). \quad (2.1)$$

Dabei ist $P(\tau \leq t)$ die Wahrscheinlichkeit dafür, dass eine Betrachtungseinheit eine Lebensdauer τ kleiner der gegebenen Zeitdauer t erreicht. Der Definitionsbereich der Ausfallwahrscheinlichkeit ist $[0, \infty)$, mit $F(0) = 0$ und $F(\infty) = 1$.

Zuverlässigkeit

Die Zuverlässigkeit ist als Komplement der Ausfallwahrscheinlichkeit eine monoton fallende Funktion:

$$R(t) = 1 - F(t) = 1 - P(\tau \leq t) = P(\tau > t). \quad (2.2)$$

Dabei ist $P(\tau > t)$ die Wahrscheinlichkeit dafür, dass eine Betrachtungseinheit eine Lebensdauer τ größer als die gegebene Zeitdauer t erreicht.

Als zentraler Begriff der Zuverlässigkeitstechnik ist diese in [24] definiert als:

„Beschaffenheit einer Einheit bezüglich ihrer Eignung, während oder nach vorgegebenen Zeitspannen bei vorgegebenen Anwendungsbedingungen die Zuverlässigkeitsforderung zu erfüllen.“

Ausfallwahrscheinlichkeitsdichte

Die Ausfallwahrscheinlichkeitsdichte oder auch Dichtefunktion $f(t)$ charakterisiert die Wahrscheinlichkeit eines Ausfalls zum Zeitpunkt t bezüglich eines Zeitintervalls der Größe dt . Die Dichtefunktion ergibt sich aus der Ableitung der Ausfallwahrscheinlichkeit:

$$f(t) = \frac{dF(t)}{dt}. \quad (2.3)$$

Ausfallrate

Die Ausfallrate $\lambda(t)$ charakterisiert das Risiko eines Bauteils auszufallen, wenn es bis zum aktuell betrachteten Zeitpunkt t überlebt hat. Somit ist sie die Wahrscheinlichkeit eines Ausfalls im Zeitintervall dt bezüglich aller intakten Einheiten zum Zeitpunkt t :

$$\lambda(t) = \frac{f(t)}{R(t)}. \quad (2.4)$$

Medianränge

Zur Auswertung von Ausfalldaten muss jedem Ausfall eine gewisse Wahrscheinlichkeit des Auftretens zugeordnet werden. Dazu müssen die Ausfalldaten aufsteigend nach ihrer Größe geordnet werden. Die Position innerhalb der nach Größe geordneten Ausfalldaten ist die Rangzahl. Über die Näherungsformel ergibt sich mit der Rangzahl die zugeordnete Ausfallwahrscheinlichkeit nach Gl. (2.5), der Medianrang $F_i(t_i)$ [25].

$$F_i(t_i) = \frac{i-0,3}{n+0,4} \quad (2.5)$$

Darin ist t_i die Ausfallzeit der i -ten Ranggröße, i die Rangzahl und n der Stichprobenumfang. Da die betrachtete Stichprobe nur einem Teil der Grundgesamtheit entspricht und die Ausfallzeiten jeder Ranggröße bei verschiedenen Stichproben zufälligen Schwankungen unterliegen, werden die zugeordneten Ausfallwahrscheinlichkeiten als Medianränge berechnet. Die Ranggröße i ist eine Zufallsgröße, die mit einer Binomialverteilung beschrieben werden kann. Aus dem Medianwert der Binomialverteilung ergibt sich der Medianrang der jeweiligen Ranggröße [26].

Zensierte Daten

Zensierte Daten treten immer dann auf, wenn ein unvollständiger Datensatz vorliegt. Dies ist in der Zuverlässigkeitstechnik der Fall, wenn aus einem Versuch Teile vor dem Versuchsende herausgenommen werden, oder wenn bei Felddaten nicht von allen Teilen die Laufzeiten bekannt sind. Prinzipiell können drei Zensierungsarten unterschieden werden: die Rechtszensierung, die Linkszensierung und die Intervallzensierung. Rechtszensierte Daten können in drei Typen eingeteilt werden, siehe Bild 2.1.

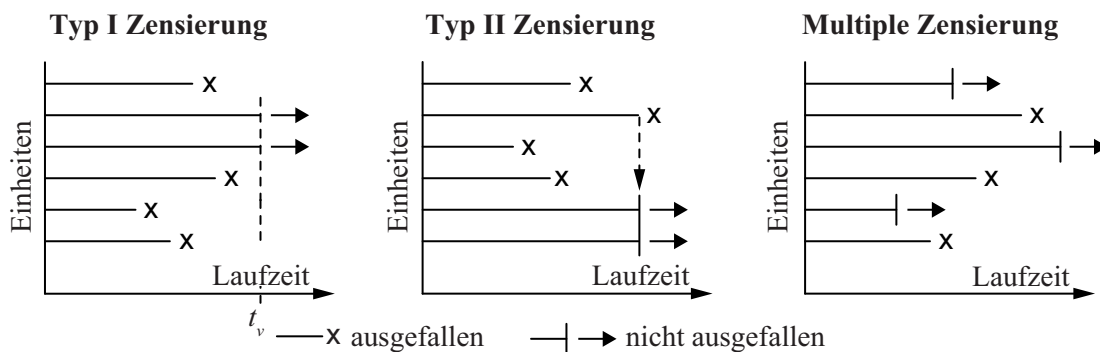


Bild 2.1: Typen von rechtszensierten Daten

Bei der Zensierung nach Typ I ist die maßgebliche Größe die maximale Laufzeit. Dabei wird ein Lebensdauerversuch nach einer vorgegebenen Zeit t_v abgebrochen. Alle bis dahin nicht ausgefallenen Einheiten sind zensiert. Bei der Zensierung nach Typ II ist die maßgebliche Größe die Anzahl an Ausfällen. Hier wird der Versuch nach einer vorgegebenen Anzahl an Versuchen abgebrochen. Für die Versuchsauswertung wird die Typ I oder Typ II zensierte Stichprobe wie eine vollständige Stichprobe behandelt. Eine multiple Zensierung liegt vor, wenn intakte Teile vor dem eigentlichen Versuchsende aus dem Versuch entfernt werden. Diese Teile waren bis zur Herausnahme aus dem Versuch intakt, es ist aber nicht bekannt ob sie zu einem späteren Zeitpunkt noch ausgefallen wären. Zur Auswertung von zensierten Daten ist eine Vielzahl an Methoden bekannt, wie beispielsweise das Anwärterverfahren oder das Johnson Verfahren, siehe [27], [28], [29].

Johnson Verfahren

Das Johnson Verfahren ist ein Verfahren zur Bestimmung der Medianränge von multipel zensierten Daten. Die Voraussetzung zur Anwendung des Verfahrens ist die Kenntnis der Laufzeiten der ausgefallenen und intakten Einheiten. Die Einheiten werden entsprechend ihrer Laufzeiten aufsteigend sortiert und die aller Einheiten erhalten eine hypothetische Rangzahl $j(t_j)$, entsprechend ihrer Laufzeit t_j innerhalb der ausgefallenen Einheiten. Die hypothetische Rangzahl berücksichtigt die zwischen den einzelnen Ausfällen liegenden intakten Einheiten. Die hypothetische Rangzahl $j(t_j)$ berechnet sich nach Gl. (2.6).

$$j(t_j) = j(t_{j-1}) + N(t_j); \quad j(0) = 0 \quad (2.6)$$

Dabei ist $j(t_{j-1})$ die vorhergehende hypothetische Rangzahl und $N(t_j)$ ein Zuwachs entsprechend Gl. (2.7).

$$N(t_j) = \frac{n+1-j(t_{j-1})}{1+(n-g)} \quad (2.7)$$

Darin ist n der gesamte Stichprobenumfang von ausgefallenen und intakten Einheiten und g die Anzahl der vor dem betrachteten Ausfall j liegenden ausgefallenen und intakten Einheiten. Mit den hypothetischen Rangzahlen $j(t_j)$ können die den jeweiligen Ausfällen zugeordneten Medianränge über Gl. (2.5) bestimmt werden.

Kaplan-Meier-Schätzer

Ein weiteres Verfahren zur Bestimmung der Wahrscheinlichkeiten von Ausfällen ist der Kaplan-Meier-Schätzer oder Produkt-Limit-Schätzer [30]. Er kann sowohl auf vollständige als auch auf zensierte Ausfalldaten angewendet werden. Es wird die Wahrscheinlichkeit geschätzt, dass ein Ereignis in einem betrachteten Zeitintervall nicht auftritt. Die geschätzte Größe ist in diesem Fall die Zuverlässigkeit $R(t)$, die Ausfallwahrscheinlichkeit lässt sich nach Gl. (2.2) als Komplement berechnen:

$$R(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} = \prod_{t_i < t} 1 - \frac{d_i}{n_i} \quad (2.8)$$

Darin ist n_i die Anzahl der zum Zeitpunkt t_i noch nicht ausgefallenen Teile und d_i die im betrachteten Zeitintervall aufgetretenen Ausfälle.

2.2 Verteilungen und Systemtheorie

Zur Darstellung des zeitlichen Verlaufs von Ausfällen dienen Verteilungsfunktionen. Mit diesen Verteilungsfunktionen oder Verteilungen wird versucht, die vorliegenden Datenpunkte, in der Zuverlässigkeitstechnik Ausfalldaten, mathematisch zu

beschreiben. In diesem Abschnitt werden die Grundlagen zu Weibullverteilungen und gemischten Verteilungen vorgestellt. Zusätzlich werden die gemischten Verteilungen in Systeme und Mischverteilungen unterteilt und deren Abgrenzung voneinander beschrieben.

2.2.1 Weibullverteilung

Eine der am häufigsten eingesetzten und vielseitigsten Verteilungsfunktionen in der Zuverlässigkeitstechnik ist die Weibullverteilung [1]. Mit ihr lassen sich unterschiedlichste Verläufe der Ausfallwahrscheinlichkeit nachbilden. Es sind fallende Ausfallraten wie bei Frühausfällen, konstante Ausfallraten wie bei Zufallsausfällen und ansteigende Ausfallraten wie bei Verschleiß- oder Alterungsausfällen darstellbar, siehe Bild 2.2. Damit können alle Bereiche der sogenannten Badewannenkurve abgedeckt werden [1].

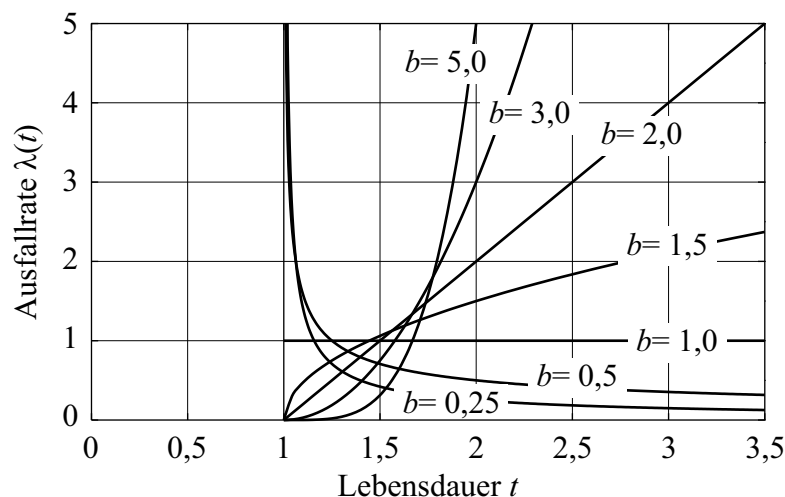


Bild 2.2: Ausfallraten $\lambda(t)$ von Weibullverteilungen mit verschiedenen Formparametern b (char. Lebensdauer $T = 2$; ausfallfreie Zeit $t_0 = 1$)

Die verschiedenen Verläufe der Verteilung sind vom Formparameter b abhängig. Der Lageparameter ist als charakteristische Lebensdauer T definiert. Als dritter Parameter kann die ausfallfreie Zeit t_0 eine Verschiebung auf der Zeitachse bewirken. Dieser Parameter tritt nur in der dreiparametrischen Form der Weibullverteilung auf. Die ausfallfreie Zeit t_0 ist im Intervall $(0, T]$ definiert. Die Verteilungsfunktion $F(t)$ und die Dichtefunktion $f(t)$ der dreiparametrischen Weibullverteilung sind gegeben durch:

$$F(t) = 1 - e^{-\left(\frac{t-t_0}{T-t_0}\right)^b} \quad (2.9)$$

$$f(t) = \frac{dF(t)}{dt} = \frac{b}{T-t_0} \left(\frac{t-t_0}{T-t_0}\right)^{b-1} e^{-\left(\frac{t-t_0}{T-t_0}\right)^b} \quad (2.10)$$

Durch Nullsetzen der ausfallfreien Zeit t_0 kann die dreiparametrische Weibullverteilung in ihre zweiparametrische Form überführt werden.

Auf Grund ihrer universellen Einsetzbarkeit und ihrer Bedeutung zur Beschreibung des Ausfallverhaltens von Maschinenbauprodukten wird im Folgenden nur die Weibullverteilung als zu schätzende Verteilungsfunktion eingesetzt.

2.2.2 Gemischte Verteilungen

Zur Beschreibung des Ausfallverhaltens von Bauteilen ist eine Vorgehensweise am sinnvollsten, bei der die aufgetretenen Ausfälle nach den jeweiligen Ausfallmechanismen separiert werden. Diese können anschließend getrennt ausgewertet werden. Dadurch kann eine Beschreibung des Ausfallverlaufs der einzelnen Ausfallmechanismen erfolgen.

Bei Felddaten können häufig verschiedene Ausfälle nicht den jeweiligen Ausfallmechanismen zugeordnet werden. Dadurch entfällt die Möglichkeit, die Ausfälle getrennt nach Ausfallmechanismen auszuwerten. Eine solche Verteilung mit mehreren Unterverteilungen ist in Bild 2.3 dargestellt.

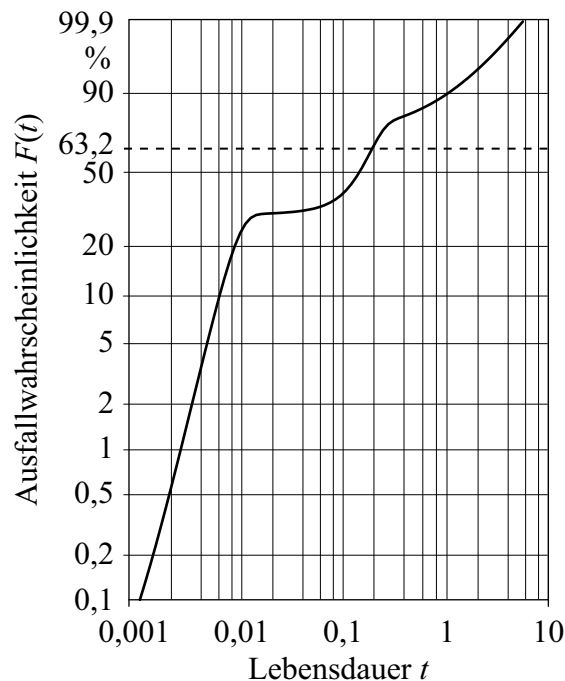


Bild 2.3: Gemischte Verteilung

Durch die doppeltlogarithmische Skalierung der Ordinate $F(t)$ und logarithmische Skalierung der Abszisse t wird der s-förmige Verlauf einer zweiparametrischen Weibullverteilung in eine Gerade transformiert. Die so entstehende Skalierung der Achsen wird im sogenannten Weibullpapier, wie in Bild 2.3 dargestellt, verwendet.

Der Verlauf der Verteilung ist durch Knickstellen, Stufen und Sattelpunkte gekennzeichnet, was auf das Auftreten von verschiedenen Unterverteilungen hindeutet. Einem solchen Verlauf können zwei Ursachen zu Grunde liegen:

- Ein Bauteil besteht aus mehreren Komponenten, die in einer Reihen- oder Parallelschaltung angeordnet sind und ein unterschiedliches Ausfallverhalten aufweisen. → boolesches Modell
- In der betrachteten Stichprobe sind zwei Populationen gemischt, die durch verschiedene Verteilungen gekennzeichnet sind, beispielsweise rote und weiße Kugeln mit unterschiedlich verteilten Durchmessern. → Mischverteilung

Boolesches Modell

Eine Kombination verschiedener Komponenten zu einem System lässt sich mit Hilfe des booleschen Modells beschreiben. Zur Darstellung von Bauteilen mit mehreren Komponenten dient das Zuverlässigkeitsblockdiagramm. Damit können Serienstrukturen bzw. Reihenschaltungen, parallele Strukturen und gemischte Strukturen aus Komponenten (K_x) mit ihren jeweiligen Verbindungen graphisch dargestellt werden, siehe Bild 2.4.

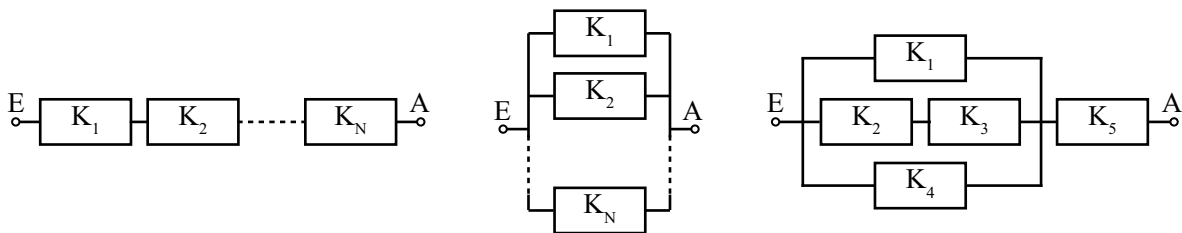


Bild 2.4: Zuverlässigkeitsblockdiagramme von Serien- (links), Parallelstruktur (Mitte) und gemischter Struktur (rechts)

Um diese Systeme über die boolesche Algebra beschreiben zu können, müssen verschiedene Anforderungen erfüllt sein [1]:

- Das System und die einzelnen Systemkomponenten können durch zwei Zustände, „intakt“ und „ausgefallen“, beschrieben werden.
- Die einzelnen Systemkomponenten sind unabhängig voneinander und beeinflussen sich nicht gegenseitig in ihrem Ausfallverhalten.
- Die Lebensdauer des Systems ist nach dem ersten Systemausfall beendet. Es ist ein nichtreparierbares System. Ist das System reparierbar, so kann nur bis zum ersten Systemausfall gerechnet werden.

Die Verteilungsfunktionen $F_S(t)$ können über die boolesche Algebra beschrieben werden. Für Seriensysteme ergibt sich die Verteilungsfunktion als das Komplement aus den Produkten der einzelnen Komplemente der Verteilungsfunktionen der Komponenten, siehe Gl. (2.11).

$$F_S(t) = 1 - \prod_{i=1}^N (1 - F_i(t)) \quad (2.11)$$

Die Ausfallwahrscheinlichkeit eines Seriensystems steigt mit zunehmender Anzahl an Einzelkomponenten an.

Für Parallelsysteme ergibt sich die Verteilungsfunktion $F_S(t)$ aus dem Produkt der Verteilungsfunktionen der Komponenten:

$$F_S(t) = \prod_{i=1}^N F_i(t). \quad (2.12)$$

Die Ausfallwahrscheinlichkeit eines Parallelsystems sinkt mit zunehmender Anzahl an Einzelkomponenten. Die Anzahl der Einzelkomponenten N eines Parallelsystems wird auch Redundanzgrad genannt.

Durch die Verknüpfung der beiden Grundtypen Serien- und Parallelsystem können die Verteilungsfunktionen von gemischten Strukturen bestimmt werden.

In Bild 2.5 ist die Verknüpfung von zwei Komponenten mit verschiedenen Ausfallverteilungen als Serien- und als Parallelsystem dargestellt. Deutlich zu erkennen ist die Zunahme der Steigung der Systemkurve des Seriensystems und die Abnahme der Steigung des Parallelsystems über der Lebensdauer t . Im Falle des Seriensystems ergibt sich eine charakteristische Lebensdauer $T_S \approx T_1$ und im Falle des Parallelsystems $T_S \approx T_2$.

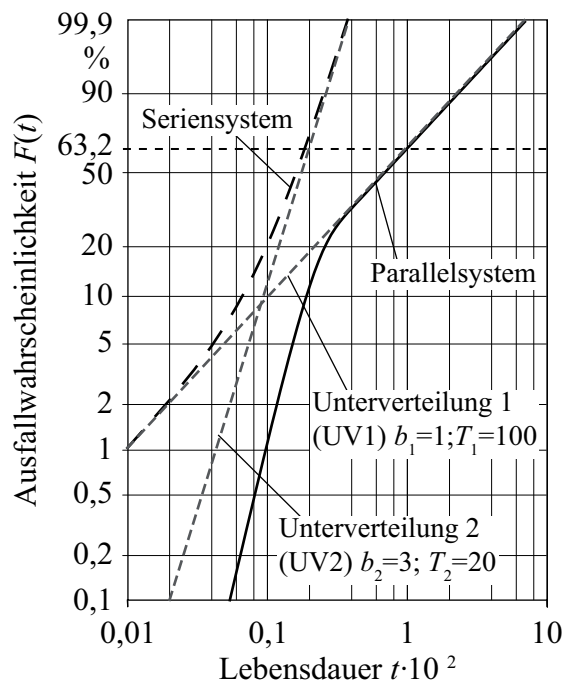


Bild 2.5: Systemausfallwahrscheinlichkeit eines Seriensystems und eines Parallelsystems mit denselben Unterverteilungen

Neben dem booleschen Modell gibt es weitere Beschreibungsmöglichkeit für die Zuverlässigkeit von Systemen, beispielsweise das Markov-Modell [31]. Markov-Modelle sind jedoch auf die Exponentialverteilung beschränkt [32], wodurch eine Anwendung im Rahmen dieser Arbeit ausgeschlossen wurde.

Mischverteilungen

Eine Mischverteilung tritt auf, wenn der betrachtete Datensatz aus mehreren Unterpopulationen besteht. Dies kann beispielsweise der Fall sein, wenn die betrachteten Einheiten aus verschiedenen Materialchargen gefertigt wurden oder eine Durchmischung von Daten mit verschiedenen zugrunde liegenden Ausfallmechanismen stattgefunden hat [33].

Die Ausfallverteilung der Mischverteilung $F_M(t)$ lässt sich damit als Summe der gewichteten Unterverteilungen $F_i(t)$ darstellen, siehe Gl. (2.13).

$$F_M(t) = \sum_{i=1}^N F_i(t) p_i \quad (2.13)$$

Dabei ist die Summe aller Gewichtungsfaktoren p_i eins. Die Anzahl der Unterpopulationen bzw. Unterverteilungen ist N . Die Dichtefunktion der Mischverteilung $f_M(t)$ ist damit:

$$f_M(t) = \sum_{i=1}^N f_i(t) p_i \quad (2.14)$$

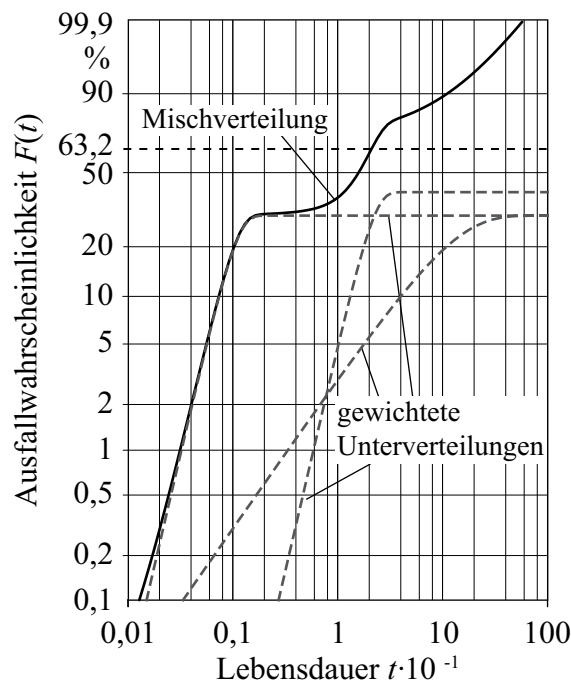


Bild 2.6: Mischverteilung aus drei Unterverteilungen mit gewichteten Unterverteilungen

In Bild 2.6 ist eine Mischverteilung mit drei Unterverteilungen dargestellt. Zusätzlich sind die gewichteten Unterverteilungen abgebildet.

Die Separierbarkeit von Mischverteilungen aus zwei Weibullverteilungen wurde von Jiang und Kececioglu in [9] untersucht. Einfach separierbare zweiparametrische Weibull-Mischverteilungen entstehen durch

- Beibehalten der Formparameter b_1 und b_2 und Festlegen der charakteristischen Lebensdauern $T_1 \ll T_2$ oder $T_1 \gg T_2$,
- Beibehalten von $T_1 < T_2$ oder $T_1 > T_2$ und Erhöhen von b_1 und b_2 .

2.2.3 Abgrenzung von gemischten Verteilungen

Aus dem Verlauf einer Verteilung aus mehreren Unterverteilungen kann auf die Anzahl der zu Grunde liegenden Unterverteilungen geschlossen werden. Jeder Knick kann als eine Änderung des Ausfallmechanismus bzw. eine Änderung der maßgeblichen Unterverteilung interpretiert werden. Dies gilt jedoch nur wenn die verschiedenen Verteilungen separierbar sind. Bild 2.7 zeigt drei Varianten für boolesche Seriensysteme und Mischverteilungen, die nicht separiert werden können.

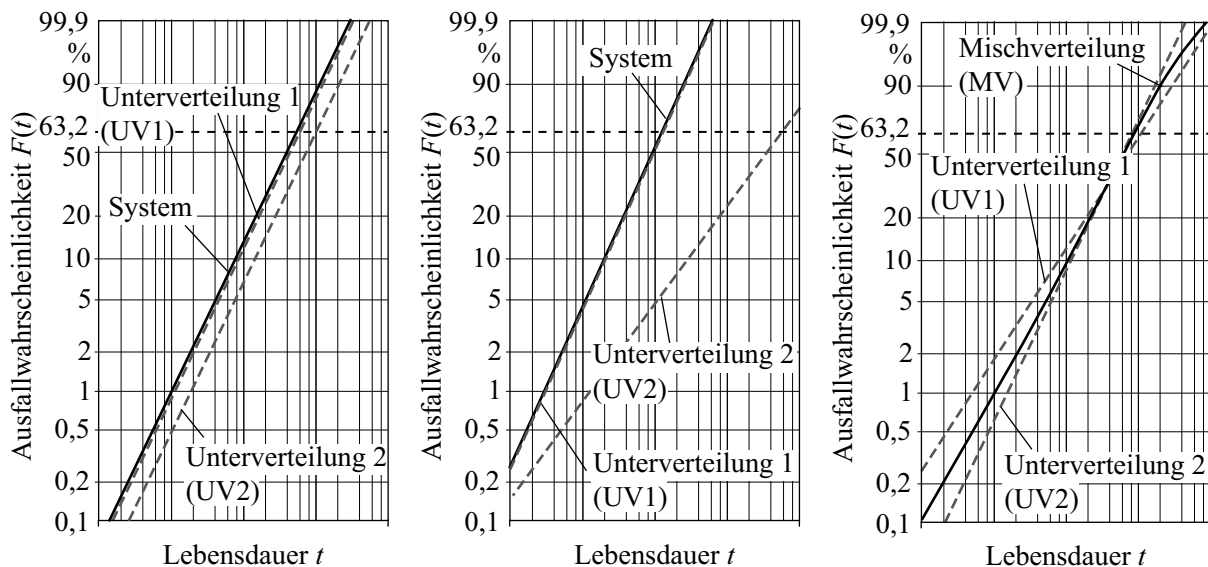


Bild 2.7: System mit parallelen Ausfallverteilungen (links), System mit sehr ungleichen char. Lebensdauern (Mitte), Mischverteilung mit aufeinanderliegenden Ausfallverteilungen (rechts)

Im Bild 2.7 links ist ein boolesches Seriensystem dargestellt. Es besteht aus zwei Komponenten, die einen gleichen Formparameter besitzen, jedoch unterschiedliche charakteristische Lebensdauern. Dadurch verlaufen beide Verteilungen parallel und der Gesamtverlauf wird hauptsächlich durch die Verteilung der Komponente mit der kleineren charakteristischen Lebensdauer bestimmt. Die Verteilungen können somit nicht separiert werden.

Im mittleren Bild in Bild 2.7 ist ein System mit verschiedenen Formparametern dargestellt. Die zweite Unterverteilung (UV2) besitzt einen Lageparameter $T_2 \gg T_1$. Dies bewirkt, dass kein Schnittpunkt der beiden Verteilungen innerhalb des dargestellten Bereichs auftritt. Somit kann die zweite Verteilung nicht identifiziert werden.

Im Bild 2.7 rechts ist eine Mischverteilung mit zwei nahezu identischen Form- und Lageparametern dargestellt. Dadurch tritt nur bei sehr hohen Ausfallwahrscheinlichkeiten eine leichte Knickstelle im Verlauf der Mischverteilungskurve auf. Dieser leichte Knick wird innerhalb der Ausfallzeiten nicht erkennbar sein. Dadurch ist eine Mischverteilung hier sehr wahrscheinlich nicht feststellbar.

Bei Felddaten kommt erschwerend hinzu, dass die Datensätze häufig sehr klein sind und durch verschiedene Einflüsse Knickstellen, Stufen und Sattelpunkte im Verlauf der Datenpunkte auftreten können. Das Auftreten von Knicken im Verlauf ist vor allem bei kleinen Stichproben nicht immer ein Zeichen einer gemischten Verteilung. Rückläufer aus dem Feld, die bei planmäßigen Wartungen getauscht werden, können, wenn sie zum gleichen Zeitpunkt getauscht werden, eine Stufe im Ausfallverlauf erzeugen, siehe Bild 2.8. Zudem können auch stochastische Effekte wie beispielsweise Ausreißer abknickende Verläufe verursachen.

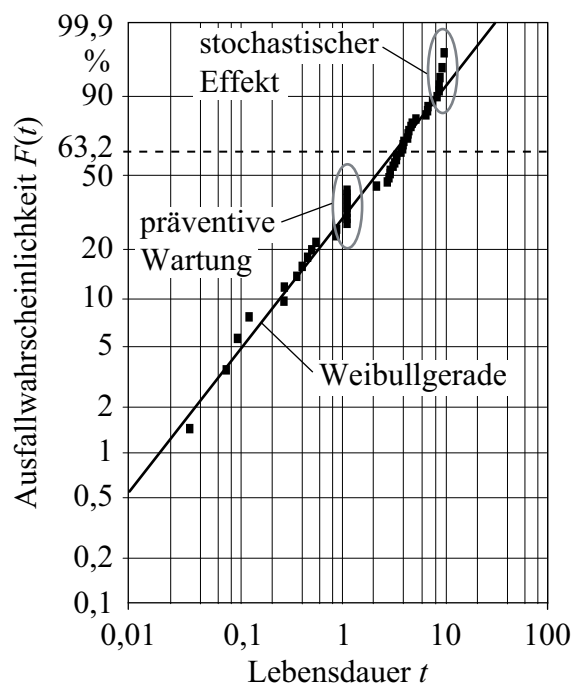


Bild 2.8: Fehlinterventionen

Um eine Aussage über die den gemischten Verteilungen zu Grunde liegenden Unterverteilungen machen zu können, ist die Kenntnis der untersuchten Daten unerlässlich. Nur wenn klare Indizien für eine gemischte Verteilung vorliegen, kann eine Auswertung der Daten als gemischte Verteilung sinnvolle Ergebnisse zeigen. Zumindest eines

der folgenden Kriterien sollte erfüllt sein, um eine Auswertung mit gemischten Verteilungen vorzunehmen:

- Es ist bekannt, dass die betrachteten Daten aus einem System stammen und es wurden mehrere Ausfallmechanismen festgestellt.
- Die Ausfalldaten stammen bekanntermaßen aus mehreren verschiedenen Stichproben.
- Eine Beschreibung des Ausfallverlaufs mit einer Verteilung führt zu einer unbefriedigenden Korrelation mit den Ausfalldaten.
- Knicke und Stufen im Verlauf deuten eindeutig auf eine gemischte Verteilung hin.
- Es ist eine physikalische Erklärung vorhanden, die auf das Auftreten mehrerer Ausfallmechanismen hindeutet.
- Die Stichprobe sollte eine gewisse Größe haben, um keine stochastischen Effekte fälschlicherweise als Knick oder Stufe zu identifizieren.

2.3 Parameterschätz- und Optimierungsverfahren

Um eine Verteilungsfunktion, wie in Bild 2.8 dargestellt, zu erhalten, müssen aus vorhandenen Ausfalldaten oder Schadensstatistiken die Parameter der Verteilungsfunktion bestimmt werden. Die dazu verwendeten Methoden werden im Folgenden für Einzelverteilungen und Mischverteilungen beschrieben.

Ebenfalls werden die Parameterschätzmethoden zur Schätzung von unvollständigen Datensätzen, sogenannten zensierten Daten, vorgestellt. Den Schwerpunkt bilden die verschiedenen Methoden zur Schätzung von Mischverteilungen. Diese werden im Einzelnen in den jeweiligen drei Hauptgruppen der graphischen, analytischen und numerischen Verfahren beschrieben.

2.3.1 Parameterschätzung von Einzelverteilungen

Zur Schätzung von Verteilungsparametern von Einzelverteilungen sind viele Methoden bekannt. Eine Übersicht und Bewertung von verschiedenen Parameterschätzmethoden findet sich in [34]. An dieser Stelle werden nur die analytischen Methoden der Regressionsanalyse und der Maximum-Likelihood Methode (MLE) vorgestellt.

Regressionsanalyse

Bei der Regressionsanalyse oder Methode der kleinsten Fehlerquadrate wird eine Ausgleichsgerade durch die Ausfalldaten gelegt [35]. Die Ausfalldaten werden als Wertepaare aus dem Ausfallzeitpunkt t_i und der zugeordneten Ausfallwahrscheinlichkeit $F(t_i)$ dargestellt. Es werden die Parameter der Ausgleichsgeraden so bestimmt, dass sich ein Minimum der Abstandsquadrate zwischen Ausfalldatenpunkten und der

Ausgleichsgeraden ergibt. Für die Ausgleichsgerade werden die Geradensteigung m und der Achsabschnitt y_a ermittelt.

$$\rho^2 = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y(F_i) - f(x(t_i)))^2 = \sum_{i=1}^n (y_i - mx_i - y_a)^2 \rightarrow \min \quad (2.15)$$

Zur Minimierung der Abstandsquadrate ρ^2 werden die partiellen Ableitungen von Gl. (2.15) nach m (Gl. (2.16)) und y_a (Gl. (2.17)) bestimmt.

$$\frac{\partial \rho^2}{\partial m} = -\sum_{i=1}^n 2(y_i - mx_i - y_a) = 0 \Rightarrow \sum_{i=1}^n (y_i - mx_i - y_a) = 0 \quad (2.16)$$

$$\frac{\partial \rho^2}{\partial y_a} = -\sum_{i=1}^n 2x_i(y_i - mx_i - y_a) = 0 \Rightarrow \sum_{i=1}^n x_i(y_i - mx_i - y_a) = 0 \quad (2.17)$$

Aus diesen Ableitungen lassen sich zwei Verteilungsparameter berechnen, beispielsweise der Formparameter b und die charakteristische Lebensdauer T der Weibullverteilung.

Maximum-Likelihood Methode

Die Maximum-Likelihood Methode ist eine häufig in der Zuverlässigkeitstechnik angewandte Methode. Die Parameterschätzung erfolgt durch die Maximierung der Likelihoodfunktion, die eine Verteilung der Zufallsvariablen der Ausfalldaten t_i als eine Funktion der Verteilungsparameter ψ_j darstellt. Für die Weibullverteilung sind die Parameter ψ_j gegeben durch die Parameter b , T und t_0 . Die Wahrscheinlichkeit L , dass eine gegebene Stichprobe optimal approximiert wird, ist am höchsten, wenn das Produkt aus den Wahrscheinlichkeiten zu den Zeitpunkten t_i im Intervall dt maximal wird. Für eine sehr große Stichprobe vom Umfang n wird das Intervall dt sehr klein.

$$L(t_1, \dots, t_n; \psi) = f(t_1; \psi) f(t_2; \psi) \dots f(t_n; \psi) = \prod_{i=1}^n f(t_i; \psi) \quad (2.18)$$

Das Maximum der Likelihoodfunktion ist für ψ_{max} innerhalb des Parameterraums von ψ als die Nullstelle der ersten Ableitung von

$$\frac{\partial L}{\partial \psi} = 0, \quad (2.19)$$

für die gilt:

$$\frac{\partial^2 L}{\partial \psi^2} < 0. \quad (2.20)$$

Das globale Maximum ψ_{max} wird aus den lokalen Maxima der zweiten Ableitung von L nach ψ bestimmt.

Zur Erleichterung der Berechnung wird die Likelihoodfunktion häufig logarithmiert. Dies ist zulässig, da der natürliche Logarithmus eine monotone Funktion ist. Durch die Logarithmierung wird aus dem Produkt in Gl. (2.18) eine Summe.

$$\ln(L(t_1, \dots, t_n; \psi)) = \ln(f(t_1; \psi)) + \ln(f(t_2; \psi)) \dots \ln(f(t_n; \psi)) = \sum_{i=1}^n \ln(f(t_i; \psi)) \quad (2.21)$$

Zur Lösung der oft nichtlinearen Ableitungen der Likelihoodfunktion kommen verschiedene Verfahren wie beispielsweise das Newton-Raphson-Verfahren [36] zum Einsatz.

2.3.2 Graphische Methoden zur Parameterschätzung von Mischverteilungen

In der Literatur finden sich eine Vielzahl von Methoden zur graphischen Analyse und zur Schätzung von Mischverteilungen und ihren Parametern [37]. Bekannte Methoden für Normalverteilungen sind die Methode nach Nothnagel [38], die Methode nach Harding [39] und die Methode nach Cassie [40]. Für die Schätzung von Weibull-Mischverteilungen werden im Folgenden die Verfahren nach Kao-Cran [5], [6] und Jensen-Petersen vorgestellt [7]. Der Vorteil von graphischen Verfahren ist im Allgemeinen der geringe Aufwand bei häufig ausreichender Genauigkeit der Ergebnisse.

Die Grundlage für die graphischen Verfahren sind Eigenschaften von sehr gut separierbaren Mischverteilungen. Dies sind:

- Die Steigung der Verteilungsfunktion der Mischverteilung strebt in einem Bereich gegen Null \rightarrow Sattelpunkt
- Die Ausfallwahrscheinlichkeit für $t = T_1 < T_2$ ist

$$F(T_1) \rightarrow 0,632p_1 \quad (2.22)$$

und für $t = T_2$

$$F(T_2) \rightarrow p_1 + 0,632(1 - p_1); \quad (2.23)$$

damit lässt sich bei bekanntem p_1 , T_1 direkt der Abszissenwert der Mischverteilung bei $0,632p_1$ ablesen.

- Die Tangenten am Verlauf der Mischverteilung für $F(t) \rightarrow 0$ und $F(t) \rightarrow 1$ sind asymptotisch parallel zu den Verläufen der Unterverteilungen

Methode nach Kao-Cran

Die Methode wurde von Kao in [5] und Cran in [6] beschrieben. Das Hauptaugenmerk bei dieser Methode liegt auf der Bestimmung des Mischanteils p_1 der ersten Unterverteilung.

Bei der Parameterschätzung werden nacheinander nachfolgend aufgeführte Schritte durchgeführt:

1. Eintragen der Ausfalldaten in ein Weibullwahrscheinlichkeitspapier.
2. Bestimmung der Tangenten $G_1(t)$ und $G_2(t)$ an den Enden der Verteilung für $F(t) \rightarrow 0$ und $F(t) \rightarrow 1$.
3. Am Schnittpunkt von $G_2(t)$ mit dem oberen Rand des Weibullpapiers wird ein Lot L gefällt. Am Schnittpunkt des Lotes L mit der Tangente $G_1(t)$ ergibt sich der Ordinatenwert als der Schätzwert für p_1 .
4. Eine eventuell vorhandene ausfallfreie Zeit t_0 wird über die Schnittpunkte mit der Abszissenachse bestimmt.
5. Die Ausfalldatenpunkte werden den beiden Unterverteilungen zugeordnet. Werte, die nahe oder auf $G_1(t)$ liegen, werden Unterverteilung eins zugeschlagen, Werte, die nahe oder auf $G_2(t)$ liegen, werden der zweiten Unterverteilung zugeschlagen.
6. Aus den nun getrennten Datensätzen werden die Parameter der einzelnen Unterverteilungen $F_1(t)$ und $F_2(t)$ bestimmt.

Bild 2.9 zeigt eine beispielhafte Auswertung mit der Methode nach Kao und Cran.

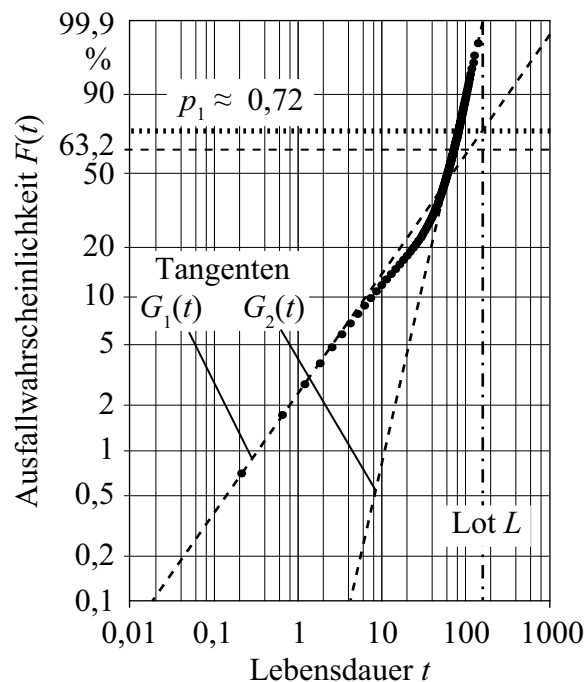


Bild 2.9: Graphische Parameterschätzmethode nach Kao und Cran

Die Anwendbarkeit dieser Methode ist auf Mischverteilungen beschränkt bei denen sich die Tangente $G_1(t)$ und das Lot L schneiden. Zudem ist die Methode sehr stark von der Art des Weibullpapiers abhängig, da die Begrenzung des Papiers eine große Rolle bei der Schätzung von p_1 spielt.

Methode nach Jensen-Petersen

Die Methode nach Jensen und Petersen [7] zielt wie die Kao-Cran Methode auf die Bestimmung der Gewichtungsfaktoren p_i ab. Der Ablauf zur Parameterschätzung ist folgendermaßen:

1. Eintragen der Ausfalldaten in ein Weibullwahrscheinlichkeitspapier und Eintragen einer Näherungskurve durch die Ausfallpunkte.
2. Bestimmung der Stelle der Mischverteilung mit der geringsten Steigung. Der Ordinatenwert der Stelle stellt den Gewichtungsfaktor p_1 für die linke Unterverteilung dar.
3. Eintragung der horizontalen Linien $H_1 = 0,632p_1$ und $H_2 = p_1 + 0,632(1-p_1)$. Die Lote L_1 und L_2 von den Schnittpunkten von H_1 bzw. H_2 mit der Mischverteilungsfunktion auf die Abszissenachse ergeben die charakteristischen Lebensdauern T_1 und T_2 .
4. Bestimmung der Formparameter b_1 und b_2 aus den Steigungen der Tangenten $G_1(t)$ und $G_2(t)$ an den Enden der Verteilung für $F(t) \rightarrow 0$ und $F(t) \rightarrow 1$.

In Bild 2.10 ist der Ablauf bei der Parameterschätzung an einem Beispiel dargestellt.

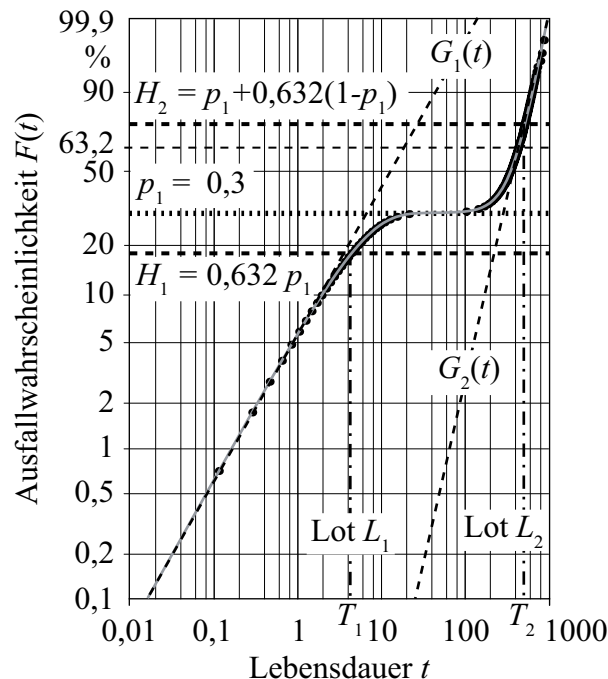


Bild 2.10: Graphische Parameterschätzungsmethode nach Jensen und Petersen

Für Mischverteilungen, die gut separierbar sind und einen Sattelpunkt beim Übergang von der ersten zur zweiten Unterverteilung aufweisen, liefert die Methode nach Jensen und Petersen gute Ergebnisse. Problematisch sind Mischverteilungen, bei denen kein Sattelpunkt vorhanden ist. Dadurch können die Stelle mit der geringsten Steigung und damit die Gewichtungsfaktoren nur schwer bestimmt werden.

2.3.3 Analytische Methoden zur Parameterschätzung von Mischverteilungen

Zur Parameterschätzung von Mischverteilungen sind auch eine Reihe analytischer Methoden bekannt [41]. Diese verbessern die eher subjektive Art der Parameterschätzung der graphischen Methoden hin zu reproduzierbaren, objektiven Ergebnissen. Von Falls wurde in [42] eine Methode zur Schätzung von zweiparametrischen Weibullverteilungen auf Basis der Momentenmethode vorgestellt. In [43] wurde von Cheng und Fu eine Methode auf Basis einer gewichteten Regressionsmethode vorgeschlagen. Im Folgenden soll auf die Parameterschätzung mittels Maximum-Likelihood Methode mit Expectation-Maximization näher eingegangen werden [44].

Maximum-Likelihood Methode mit Expectation-Maximization Algorithmus (MLE/EM)

In Kapitel 2.3.1 wurde bereits die Maximum-Likelihood Methode für einzelne Verteilungen beschrieben. Für eine einzelne Weibullverteilung müssen in der dreiparametrischen Form die Parameter b , T und t_0 geschätzt werden. Beim Übergang auf eine Mischverteilung aus zwei Unterverteilungen mit dreiparametrischen Weibullverteilungen müssen insgesamt sieben Parameter, b_1 , T_1 , t_{01} , b_2 , T_2 , t_{02} und der Gewichtungsfaktor p , berechnet werden. Bei zwei Unterverteilungen reicht der Parameter p aus, da sich der zweite Gewichtungsfaktor über das Komplement $1-p$ bestimmen lässt. Für die logarithmierte Likelihoodfunktion der Mischverteilung mit N Unterverteilungen ergibt sich

$$\ln(L(t_i | \psi, p_k)) = \sum_{i=1}^n \ln(f_M(t_i | \psi)) - \lambda \left(\sum_{k=1}^N p_k - 1 \right). \quad (2.24)$$

Darin ist ψ der Parametervektor der Verteilungsparameter aller Unterverteilungen N mit den Parametern ψ_k . Der Lagrange-Multiplikator berücksichtigt die Nebenbedingung der Optimierung, dass die Summe aller $p_k = 1$ ist.

Zur Maximierung der Likelihoodfunktion werden die partiellen Ableitungen der Parameter (Gl. (2.25)) und der Gewichtungsfaktoren (Gl. (2.26)) gebildet.

$$\frac{\partial \ln L}{\partial \psi_k} = \sum_{i=1}^n \frac{p_k}{\partial f_M(t_i | \psi)} \cdot \frac{\partial f_k(t_i | \psi_k)}{\partial \psi_k} \quad (2.25)$$

$$\frac{\partial \ln L}{\partial p_k} = \sum_{i=1}^n \frac{f_k(t_i | \psi_k)}{f_M(t_i | \psi)} - \lambda \quad (2.26)$$

Zur Berechnung des Maximums werden die Ableitungen aus Gl. (2.25) Null gesetzt. Wird Gl. (2.26) mit den einzelnen Gewichtungsfaktoren p_k multipliziert ergibt sich im Zähler die Summe der Ausfalldichte der Unterverteilung k und im Nenner die Summe der Ausfalldichte der Gesamtverteilung f_M . Aufsummiert über die k Unterverteilungen

ergibt sich der Lagrange-Multiplikator $\lambda = n$. Insgesamt ergibt sich ein Gleichungssystem aus $4N-1$ Gleichungen, je drei Weibullparameter b , T , t_0 und $N-1$ Gewichtungsfaktoren p .

Im Allgemeinen kann kein Ausfall einer Unterverteilung direkt zugeordnet werden, es liegen also unvollständige Daten vor. Es kann für die unvollständigen Daten nur iterativ versucht werden, das Maximum des Likelihoodwertes zu erreichen. Um die fehlende Zuordnung der Ausfälle zu umgehen, wird über den Expectation-Maximization-Algorithmus (EM-Algorithmus) der unvollständige Datensatz vervollständigt [41], [45]. Dazu kann mit dem Satz von Bayes eine bedingte Wahrscheinlichkeit $P_k(k|t_i)$ für jeden Ausfall i zum Zeitpunkt t_i bestimmt werden, dass er zur Unterverteilung (UV) k gehört. Allgemein ist der Satz von Bayes definiert durch:

$$P(k|t_i) = \frac{P(\text{Ausfall } i \in UV \ k \cap \text{Ausfall } i \text{ zum Zeitpunkt } t_i)}{P(\text{Ausfall } i \text{ zum Zeitpunkt } t_i)}. \quad (2.27)$$

Übertragen auf die Anwendung zur Mischverteilungsschätzung bedeutet das:

$$P(k|t_i) = \frac{p_k f_k(t_i)}{f_M(t_i)}. \quad (2.28)$$

Aus der Likelihoodfunktion und dem Satz von Bayes folgen zwei Schritte, die nacheinander ablaufen. Zuerst werden die Parameter geschätzt (Expectation-Schritt), anschließend wird mit diesen Parametern die Likelihoodfunktion maximiert (Maximization-Schritt). Dazu müssen die über Gl. (2.28) geschätzten Wahrscheinlichkeiten als Gewichtungsfaktoren in die Likelihoodfunktion eingesetzt werden.

Die Gl. (2.28) aufgelöst nach $f_k(t_i)/f_M(t_i)$ und in Gl. (2.26) eingesetzt ergibt:

$$p_k = \frac{1}{n} \sum_{i=1}^n P(k|t_i). \quad (2.29)$$

Damit ergibt sich für die Ableitungen der Likelihoodfunktion:

$$\frac{\partial \ln L}{\partial \psi_k} = \sum_{i=1}^n P_k(t_i) \frac{\partial \ln(f_k(t_i | \psi_k))}{\partial \psi_k}. \quad (2.30)$$

Zur Lösung dieses Gleichungssystems ist ein iteratives Lösungsverfahren erforderlich, beispielsweise das Nelder-Mead-Verfahren [46] oder der Levenberg-Marquardt-Algorithmus [47].

In Bild 2.11 ist ein Ablaufschema der Parameterschätzung mit dem MLE/EM-Algorithmus dargestellt.

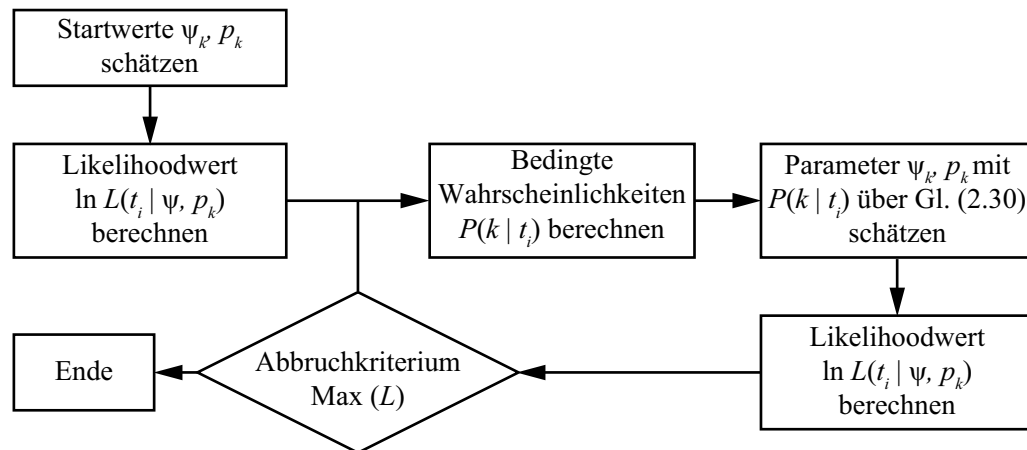


Bild 2.11: Ablauf bei der Parameterschätzung mit dem MLE/EM-Algorithmus

2.3.4 Numerische Optimierungsmethoden

Mit der Zunahme an günstiger und leicht verfügbarer Rechenkapazität nehmen auch die Möglichkeiten und die Anwendbarkeit von numerischen Methoden zu. Diese Methoden zeichnen sich durch ihren im Grundansatz einfachen Aufbau und ihren meist hohen Rechenaufwand aus.

Optimierungsmethoden lassen sich nach dem Anwendungsbereich in lineare Optimierung und nichtlineare Optimierung einteilen. Bei linearen Optimierungsproblemen ist die Zielfunktion linear. Für die Parameterschätzung von Mischverteilungen sind wegen der Nichtlinearität der Zielfunktion die in diesem Abschnitt vorgestellten nichtlinearen Optimierungsmethoden anwendbar. Das in Bild 2.12 dargestellte Schema zeigt die Zuordnung der verschiedenen nichtlinearen Optimierungsmethoden.

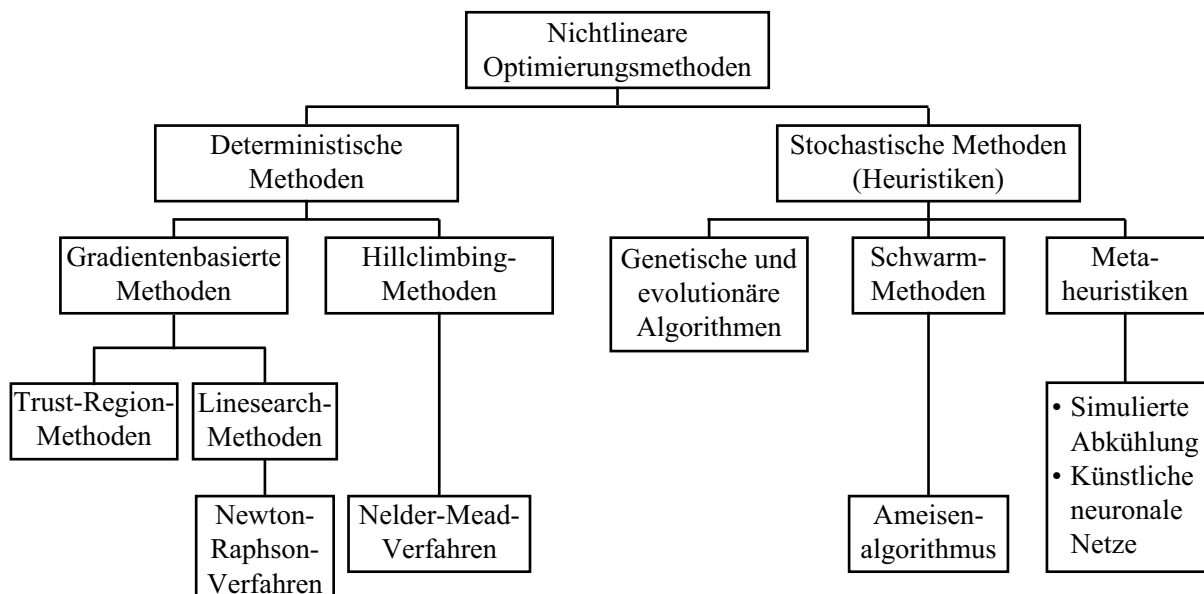


Bild 2.12: Übersicht und Einordnung von Optimierungsmethoden

In diesem Abschnitt werden verschiedene Methoden der nichtlinearen Optimierung vorgestellt, die sich für die Parameterschätzung von Mischverteilungen eignen. Die Übersichten und kurzen Beschreibungen sind angelehnt an [48]. Es soll hiermit lediglich ein Überblick über mögliche Methoden zur Anwendung auf die Parameterschätzung von Mischverteilungen gegeben werden.

Nichtlineare Optimierungsmethoden lassen sich in deterministisch und stochastisch arbeitende Methoden unterscheiden. Deterministische Methoden ergeben bei wiederholter Nutzung und unveränderten Anfangsbedingungen dasselbe Ergebnis, da sie einem festgelegten Schema folgen. Stochastische Methoden oder Heuristiken sind meist iterative Methoden, die, aufbauend auf vorhandenen Informationen, versuchen, in begrenzter Zeit eine möglichst gute Lösung zu finden. Dabei ist die gefundene Lösung nicht zwangsläufig die global beste [8].

Trust-Region-Methoden

Die Grundlagen der Trust-Region-Methoden wurden von Powell in [49] eingeführt und von Moré in [50] erweitert. Sie basieren auf einem vereinfachten lokalen Modell in der Umgebung eines Suchpunktes. Dabei werden komplexe Funktionen in der Umgebung des Suchpunktes linearisiert. Die Größe des lokalen Modells richtet sich nach dem Trust-Region-Radius. Das Modell wird aus dem lokalen Gradienten der betrachteten Funktion gebildet. Dadurch ist es möglich, schrittweise durch die Lösung eines vereinfachten Problems zu einem Optimum zu kommen. Die Trust-Region Methode wurde von Bracke in [51] zur Mischverteilungsanalyse im WCF-Ansatz (Weighted Combined Function) eingesetzt.

Linesearch-Methoden

Linesearch-Methoden sind gradientenbasierte Methoden, die schrittweise den Zielwert eines Suchpunktes verbessern. Festgelegt werden müssen die Abstiegsrichtung und die Schrittweite im betrachteten Iterationsschritt. Der Einsatzbereich ist auf die Optimierung von stetigen Zielfunktionen beschränkt. Linesearch-Methoden sind sehr stark vom Startwert abhängig, ein ungünstig gewählter Startwert führt häufig zur Konvergenz gegen ein lokales Optimum. In bestimmten Fällen kann es zu oszillierenden Ergebnissen der Zielfunktion für die jeweiligen Iterationsschritte kommen. Dies tritt auf, wenn sich zwei Lösungspunkte abwechseln. Der Vorteil der Methoden ist eine rasche Konvergenz hin zu einer Lösung. Die bekannteste Linesearch-Methode ist das Newton-Raphson-Verfahren [52].

Hillclimbing-Methoden

Hillclimbing-Methoden sind robuste Methoden, die keine Ableitung der Zielfunktion benötigen. Eine bekannte Methode aus dem Bereich der Hillclimbing-Methoden ist

das Nelder-Mead-Verfahren, auch bekannt als Downhill-Simplex-Methode. Bei dieser Methode werden für einen N-dimensionalen Parameterraum $N+1$ Punkte ausgewählt [52]. Diese Punkte werden als Eckpunkte eines Simplex verwendet. Für ein eindimensionales Problem ist der Simplex eine Strecke, im Zweidimensionalen ein Dreieck und im dreidimensionalen Parameterraum ein Tetraeder. Bei jedem Iterationsschritt wird die Ecke des Simplex mit dem schlechtesten Zielfunktionswert durch einen anderen Wert ersetzt. Diese Methode ist sehr robust, kann aber auch zu einem Nebenoptimum konvergieren.

Genetische und evolutionäre Algorithmen

Genetische und evolutionäre Algorithmen gehören zu den Heuristiken. Diese Methode ist Gegenstand des nachfolgenden Kapitels 3.1, in dem die Grundlagen dieser Methode eingehender erläutert werden.

Schwarm-Methoden

Wie die genetischen Algorithmen zählen auch die Schwarm-Algorithmen zu den Heuristiken. Diese Methoden sind, wie die genetischen Algorithmen, von der Natur inspiriert. Bekannte Vertreter dieser Gruppe von Methoden sind die Schwarm- und Ameisenalgorithmen [53], [54]. Grundlage des Ameisenalgorithmus ist eine Vielzahl an voneinander unabhängigen Agenten. Im Verlauf der Optimierung verhält sich eine Mehrheit der Agenten nach und nach ähnlich, was auf ein Optimum schließen lässt. Auch bei Ameisenalgorithmen ist das Auffinden eines globalen Maximums nicht garantiert. Ein Beispiel für die Anwendung des Ameisenalgorithmus in der Zuverlässigkeitstechnik geben Agarwal und Sharma in [55].

Metaheuristiken

Metaheuristiken sind Methoden, die lokale Optimierungsfunktionen und globale Suchstrategien miteinander kombinieren. Dadurch ist es möglich, schlechten lokalen Optima zu entgehen und näher an das globale Optimum zu gelangen.

Ein wichtiger Algorithmus aus dem Bereich der Metaheuristiken ist die simulierte Abkühlung. Sie basiert auf dem physikalischen Prinzip, dass bei der langsamen Abkühlung eines Metalls die Atome ausreichend Zeit haben, sich im Kristallgitter anzuordnen. Dadurch wird ein energiearmer Zustand des Gitters und somit ein Optimum erreicht. Der Algorithmus arbeitet iterativ. Es werden in jedem Iterationsschritt zwei Lösungsmöglichkeiten verglichen, die aktuell beste Lösung und eine neu generierte. Ist die neue Lösung besser, wird sie in jedem Fall angenommen, ist sie dagegen schlechter, wird sie mit einer gewissen Wahrscheinlichkeit trotzdem akzeptiert. Dadurch ist es möglich, lokalen Optima zu entkommen. Mit zunehmender Iterationsanzahl nimmt die

Wahrscheinlichkeit, eine schlechtere Lösung als die aktuell beste Lösung zu akzeptieren, ab. Dadurch wird eine Konvergenz des Algorithmus erreicht.

Eine weitere Methode aus dem Bereich der Metaheuristiken sind künstliche neuronale Netze. Diese Methode wird in Kapitel 3.2 näher beschrieben.

Zusammenfassung zu den numerischen Methoden

Alle hier vorgestellten Methoden sind prinzipiell zur Parameterschätzung von Mischverteilungen geeignet. Entscheidend ist jeweils die Definition der Zielfunktion, die für jeden Ansatz speziell gewählt werden muss. Viele Methoden verwenden als Zielfunktion die Lösungen der Ableitungen der Likelihoodfunktion [56], [57], [58]. In dieser Arbeit werden genetische Algorithmen und künstliche neuronale Netze auf die Anwendbarkeit zur Parameterschätzung von Weibullmischverteilungen untersucht.

2.4 Monte-Carlo Methode

Die Monte-Carlo Methode ist ein Verfahren, das auf dem Prinzip der Zufallsstichprobe beruht. Der Name leitet sich vom Stadtteil Monte Carlo in Monaco ab, der durch seine Spielbank Bekanntheit erlangte.

Durch die sehr häufige Durchführung von Zufallsexperimenten können näherungsweise Lösungen komplexer analytischer Probleme gewonnen werden. Eine häufige Anwendung ist die numerische Integration komplexer Funktionen [59]. Die Wirksamkeit der Methode hängt entscheidend von den verwendeten Zufallszahlen ab [60]. Zufallszahlen können durch Hilfsmittel, wie beispielsweise einem Würfel, erzeugt werden.

Zur Erzeugung großer Mengen an Zufallszahlen werden Pseudozufallszahlengeneratoren verwendet. Da diese Pseudozufallszahlen durch einen deterministischen Algorithmus berechnet werden, sind sie nicht „zufällig“. Bei gleichen Eingangswerten werden immer dieselben Zufallszahlen erzeugt. Durch verschiedene Verfahren können jedoch Zufallszahlen hoher Güte erzeugt werden, die sich für die Verwendung mit der Monte-Carlo Methode eignen [61]. Die Hauptkriterien für die Güte von Zufallszahlen sind eine gleichverteilte Auftretenswahrscheinlichkeit und die Unabhängigkeit der Zahlen untereinander.

Zur Erzeugung beliebig verteilter Zufallszahlen dient die Inversionsmethode. Mit ihr können die gleichverteilten Zufallszahlen in eine gewünschte Verteilung überführt werden. Mit der gleichverteilten Zufallsgröße κ im Intervall $(0, 1)$ kann über Gl. (2.31) die Zufallsgröße t aus der Inversen der Verteilungsfunktion $F(t)$ gewonnen werden.

$$\kappa = F(t) \Leftrightarrow t = F^{-1}(\kappa) \quad (2.31)$$

In Bild 2.13 ist die Erzeugung beliebig verteilter Zufallszahlen mit der Inversionsmethode dargestellt.

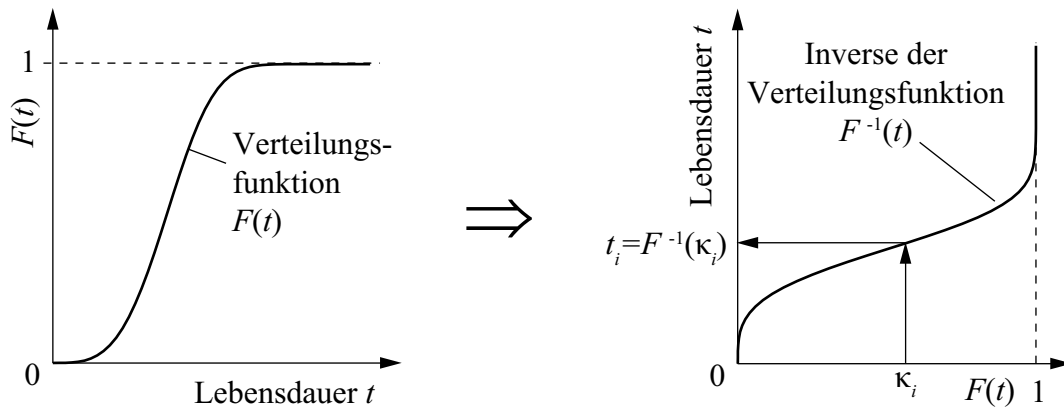


Bild 2.13: Erzeugung von beliebig verteilten Zufallszahlen t_i über die Inversionsmethode mit der Verteilungsfunktion $F(t)$

Für die Erzeugung der weibullverteilten Zufallszahlen, die in dieser Arbeit genutzt werden, ergibt sich die Inverse der Ausfallwahrscheinlichkeit der Weibullverteilung nach Gl. (2.32).

$$t_i = F^{-1}(\kappa_i) = t_0 + (T - t_0)(-\ln(1 - \kappa_i))^{\frac{1}{b}} \quad (2.32)$$

In dieser Arbeit wird die Inversionsmethode zur Erzeugung synthetischer Datensätze verwendet. Diese Datensätze werden zur Evaluation der Funktion der numerischen Methoden genutzt.

3 Grundlagen numerischer Optimierungsmethoden

In Kapitel 2.3.4 wurden verschiedene numerische Optimierungsmethoden den Bereichen der deterministischen Methoden und der stochastischen Methoden bzw. Heuristiken zugeordnet. Die in diesem Abschnitt näher beschriebenen Optimierungsverfahren genetische Algorithmen und künstliche neuronale Netze sind dem Bereich der Heuristiken zuzuordnen.

3.1 Grundlagen genetischer Algorithmen

Die genetischen Algorithmen (GA) sind ein Teilgebiet der evolutionären Algorithmen. Diese Algorithmen basieren auf dem Vorbild der natürlichen Evolution, wie sie von Darwin in [62] veröffentlicht und beschrieben wurde. Vereinfachend kann die natürliche Evolution als ein Selektionsprozess verstanden werden. Hierbei sind in einer Umgebung verschiedene Individuen einer Population vorhanden. Alle diese Individuen haben unterschiedliche Eigenschaften, was über ihren Erfolg zu überleben bzw. sich zu reproduzieren entscheidet. Je besser das Individuum an seine Umwelt angepasst ist, umso größer ist die Wahrscheinlichkeit, sich zu reproduzieren. Somit setzen sich für das Überleben günstige Eigenschaften durch. Ungünstige Eigenschaften werden dagegen nach und nach aus der Population entfernt.

3.1.1 Varianten der evolutionären Algorithmen

Auf der Grundlage dieser natürlichen Evolution entstanden verschiedene Varianten einer künstlichen Evolution zur Optimierung von mathematischen und technischen Fragestellungen. Nach [63] lassen sich die zusammenfassend als evolutionäre Algorithmen bezeichneten Methoden in vier Grundtypen aufteilen:

- Genetische Algorithmen
- Evolutionsstrategien
- Genetische Programmierung
- Evolutionäre Programmierung

All diesen Grundtypen ist ein ähnliches Vorgehen gemeinsam. Dabei wird eine Vielzahl an Lösungsmöglichkeiten für ein bekanntes Problem erzeugt. Daraus werden die geeignetsten ausgewählt und einer künstlichen Evolution unterworfen, um eine optimale Lösung des Problems zu erhalten.

Der Typ der genetischen Algorithmen wurde von Holland in [64] beschrieben, dieser Typ wird auch als kanonischer genetischer Algorithmus bezeichnet. Es werden binär

dargestellte Lösungsvarianten verwendet. Die Evolution wird hauptsächlich durch Rekombination der Lösungen realisiert.

Evolutionsstrategien wurden erstmals von Rechenberg und Schwefel [65] als Optimierungsmethode verwendet. Bei dieser Variante werden die Lösungsvarianten als reelle Zahlen dargestellt. Die Evolution wird maßgeblich durch die Mutation bestimmt.

Die genetische Programmierung geht auf eine Erweiterung der evolutionären Algorithmen durch Koza in [66] zurück. Bei diesem Typ sind die verwendeten Lösungsvarianten keine Zahlenwerte, sondern ganze Lösungsmodelle. Es werden Modelle gesucht, mit denen sich ein bestimmtes Problem möglichst gut lösen lässt.

Die evolutionäre Programmierung wurde maßgeblich durch Fogel, Owens und Walsh [67] geprägt. Bei diesem Ansatz werden nicht nur einzelne Individuen als maßgeblich für die Qualität einer Generation betrachtet, sondern die gesamte Population. Es werden wie bei den Evolutionsstrategien die Lösungsvarianten als reelle Größen dargestellt. Das zentrale Verfahren zur Generierung neuer Individuen ist die Mutation.

Eine weitere Optimierungsmethode, die teilweise den evolutionären Algorithmen zugehörig ist, sind die memetischen Algorithmen [68], [69], [70]. Bei dieser Methode werden die evolutionären Algorithmen mit einer lokalen Suchfunktion erweitert, um die Qualität der Lösungen zu erhöhen.

3.1.2 Anwendung evolutionärer Algorithmen in der Zuverlässigkeitstechnik

Die Häufigkeit der Anwendung von genetischen Algorithmen in der Zuverlässigkeitstechnik steigt mit der Zunahme an leicht verfügbarer Rechnerkapazität stark an. Vor allem im Bereich der Instandhaltung oder im Bereich der Optimierung der Zuverlässigkeit von Verteilernetzen ist der Nutzen sehr groß. Im Folgenden sollen einige Anwendungen in der Zuverlässigkeitstechnik aufgezeigt werden.

Die allgemeine Anwendung der genetischen Algorithmen zur Lösung von Optimierungsproblemen in der Zuverlässigkeitstechnik wurde von Marseguerra, Zio und Martorell in [71] untersucht. In [72] wurden von Martorell et al. genetische Algorithmen zur Optimierung von Testintervallen verwendet. Martorell et al. untersuchten in [73] die Anwendung von genetischen Algorithmen im Bereich der Wartung und Maschinenüberwachung. Aufbauend auf den Ergebnissen von Thomas et al. in [74] veröffentlichten Jenab, Kazeminia und Liu in [75] einen Ansatz zur Schätzung von Verteilungsparametern von einfachen Weibullverteilungen. Von Tolson et al. [76] wurden genetische Algorithmen zusammen mit der First-Order Reliability Method (FORM) zur Optimierung von Wasserversorgungsnetzen eingesetzt. Shayanfar, Massah und Rahami nutzten in [77] genetische Algorithmen und neuronale Netze, um

Komponenten in booleschen Systemen Verteilungsparameter zuzuordnen und so eine bestimmte Zuverlässigkeit des Gesamtsystems zu erreichen. In [78] wurde von Huang, Tian und Zuo der Redundanzgrad von Multi-State Systemen mit genetischen Algorithmen optimiert. Marseguerra, Zio und Podofillini untersuchten in [79] die Anwendung einer Kombination von genetischen Algorithmen und der Monte-Carlo Methode zur Optimierung von Systemen in Kernkraftwerken.

Die Anwendung von genetischen Algorithmen zur Schätzung von Verteilungsparametern von Einfachverteilungen in [75] lieferte gute Ergebnisse. Für diese Anwendung sind jedoch effizientere Methoden, siehe Kapitel 2.3.1, bekannt. Die Anwendung auf das komplexere Problem der Analyse von gemischten Verteilungen ist der Hauptteil dieser Arbeit. Der Vorteil der genetischen Algorithmen ist die breite Anwendbarkeit nicht nur auf die Parameterschätzung von Mischverteilungen, sondern auch von booleschen Systemen. Durch die Multiplikation der einzelnen Verteilungen bei booleschen Systemen ist die Anwendung der Maximum-Likelihood Methode schwierig, da sich die entsprechenden Ableitungen nach den einzelnen Verteilungsparametern nur schlecht bestimmen lassen. Zudem ist bei den gradientenbasierten Verfahren vor allem die Unstetigkeit an den Knickstellen bei Verteilungen mit vorhandener ausfallfreier Zeit t_0 kritisch.

3.1.3 Begriffe der genetischen Algorithmen

Genetische Algorithmen arbeiten nach dem Vorbild der natürlichen Evolution. Darum sind viele der verwendeten Begrifflichkeiten der Evolutionslehre entlehnt. Im Folgenden werden die zentralen Begriffe vorgestellt und erläutert. Einzelne wichtige Begriffe wie Fitness, Selektion oder Modifikation werden in gesonderten Abschnitten beschrieben.

Generation

Eine Generation umfasst alle Individuen innerhalb der aktuell betrachteten Iteration des genetischen Algorithmus. Die Anzahl der Generationen wird zu Beginn des Algorithmus festgelegt und ist an die Problemstellung angepasst. Jeder Iterationsschritt des genetischen Algorithmus entspricht einer Generation an Lösungsvarianten.

Individuum

Ein Individuum ist eine Lösungsvariante des zu optimierenden Problems innerhalb einer Generation. Es ist charakterisiert durch die Parameter der Kodierung. Das Individuum ist die Repräsentation einer Lösungsvariante, die alle Parameter der Lösung enthalten muss.

Kodierung

Die Kodierung ist die Entsprechung des Genotyps in der natürlichen Evolution. Der Genotyp umfasst alle genetischen Informationen eines Lebewesens, die in der DNS gespeichert sind. Die Kodierung ist demnach die Darstellung der Parameter eines Individuums. Die Anzahl der Stellen einer Kodierung ist konstant über alle Generationen und Individuen.

Allel

Das Allel bezeichnet die Ausprägung eines Gens. Im Kontext des genetischen Algorithmus ist dies der aktuelle Wert an einer einzelnen Stelle der Kodierung. Bei einer binären Kodierung können dies die Werte „0“ und „1“ sein.

Elter

Als Elter wird ein Individuum bezeichnet, das zur Erzeugung von Nachkommen mit den genetischen Operatoren Rekombination und Mutation genutzt wird.

Nachkomme

Nachkommen sind die aus der Veränderung von Eltern erzeugten resultierenden Individuen. Sie entstehen durch die Anwendung der genetischen Operatoren auf die Elternindividuen.

Population

Die Population ist die Gesamtheit aller Individuen, also Lösungsvarianten einer Generation. Sie umfasst sowohl Eltern als auch Nachkommen. Die Größe der Population ist im Allgemeinen konstant über alle Iterationen.

3.1.4 Ablauf des genetischen Algorithmus

Die Nachbildung der natürlichen Evolution durch einen genetischen Algorithmus führt zu einer iterativen Optimierung [80], [81]. Bei der natürlichen Evolution erfolgt die Auslese der Individuen nach dem Kriterium der besten Anpassung an die Umwelt. Schlechter angepasste Individuen werden nach und nach von besser angepassten verdrängt. Zur Abbildung dieses Vorgangs auf die Optimierungsmethode muss ein Bewertungskriterium eingeführt werden, nach dem die Individuen eingeteilt werden. Dieses Kriterium wird Fitnessfunktion genannt. Eine eingehende Beschreibung der Fitnessfunktion findet sich in Kapitel 3.1.5.

Die Verbesserung der jeweiligen Individuen wird in der natürlichen Evolution durch die Evolutionsfaktoren der Modifikation und der Selektion bestimmt. Dies wird in genetischen Algorithmen in ähnlicher Weise durch die Modifikationsoperatoren Rekombination und Mutation erreicht.

bination und Mutation (siehe Kapitel 3.1.7) und verschiedene Selektionsvarianten (siehe Kapitel 3.1.8) abgebildet.

Zur Nachbildung des kompletten Evolutionsprozesses müssen die Schritte der Modifikation und Selektion wiederholt schrittweise nacheinander ausgeführt werden. Ein Optimierungsschritt wird Iteration oder Iterationsschritt genannt. Dieser Ablauf ist in Bild 3.1 dargestellt.

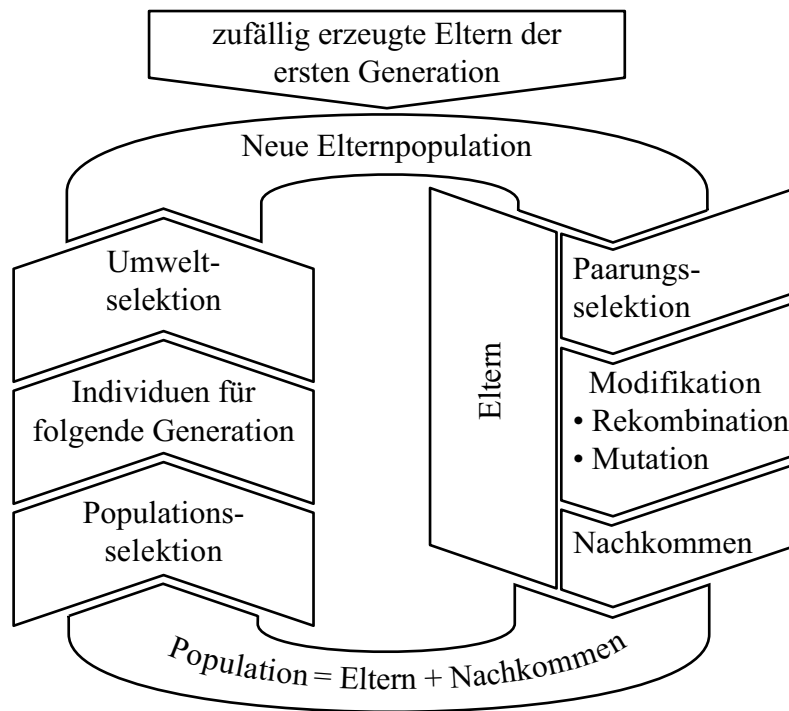


Bild 3.1: Ablauf des genetischen Algorithmus

Zu Beginn des Ablaufs wird zufällig eine erste Generation aus Eltern erstellt. Die Eltern sollten einen möglichst großen Bereich des Parameterraums des untersuchten Problems abdecken, um eine große Vielfalt in der Population zu erhalten. Die Parameter werden mittels der Kodierung für jedes erzeugte Individuum gleich dargestellt.

Die Eltern werden der ersten Selektion, der Paarungsselektion, unterworfen, um diejenigen auszuwählen, die zur Modifikation genutzt werden sollen. Anschließend werden auf die ausgewählten Eltern die Modifikationsoperatoren Rekombination und Mutation angewendet. Aus den daraus entstehenden Nachkommen und den unveränderten Elternindividuen entsteht eine neue Population. Aus dieser Population wird bei der Populationsselektion ausgewählt ob für die weiteren Schritte sowohl Eltern als auch Nachkommen oder nur Nachkommen zugelassen sind. Zusätzlich kann auch ein bestimmtes Verhältnis von Eltern zu Nachkommen innerhalb der Population definiert werden. Aus den nach der Populationsselektion verbliebenen Individuen wird anschließend durch die Umweltselektion die Elternpopulation für den nächsten Durchlauf ausgewählt. Alle

drei Selektionen sind voneinander unabhängig und werden nach den jeweiligen Erfordernissen festgelegt.

Das Ende der künstlichen Evolution wird üblicherweise nach einer festgelegten Anzahl an Generationen, also Iterationen, erreicht. Eine andere, in dieser Arbeit nicht weiter verfolgte Variante setzt als Endbedingung des Algorithmus das Erreichen eines Schwellwertes der Fitness oder einen Fitnesswert, der sich über mehrere Generationen nicht mehr verbessert.

3.1.5 Fitnessfunktion

Die Fitnessfunktion oder Zielfunktion nimmt als Bewertungsfunktion eine zentrale Rolle ein [82]. Mit ihr kann die Güte eines erzeugten Individuums bestimmt werden, was eine Vergleichbarkeit der Individuen untereinander erlaubt. Die Zielfunktion muss problemspezifisch definiert werden. Im allgemeinen Fall gilt für ein Optimierungsproblem:

$$Z_{opt} = Z(\varphi_{opt}) = \min(Z(\varphi)) \quad \text{oder} \quad Z_{opt} = Z(\varphi_{opt}) = \max(Z(\varphi)). \quad (3.1)$$

Die Zielfunktion $Z(\varphi)$ muss je nach Problemstellung minimiert oder maximiert werden. Es sind Parameter des Parameterraums φ im Definitionsbereich \mathbb{R}^v definiert, wobei v die Anzahl der Parameter ist. Für das Problem des Handlungsreisenden, siehe Bild 1.2, könnte die Zielfunktion beispielsweise die Summe der Einzelstrecken zwischen den Zielorten sein. Das Optimum wäre in diesem Fall die kleinste Summe bzw. das Minimum der Zielfunktion.

3.1.6 Kodierung

Die Kodierung der Individuen dient dazu, die Eigenschaften des Individuums darzustellen. Dazu gibt es neben der reellwertigen Kodierung auch binäre Kodierungen. Die Anwendungen mit binärer Kodierung überwiegen gegenüber denen mit reellwertiger Kodierung.

Anstelle des Standard-Binärsystems wird häufig der Gray-Code verwendet. Beim Standard-Binärsystem müssen meist mehrere Stellen der binärkodierte Zahl verändert werden, um eine benachbarte Zahl im Dezimalsystem darzustellen. Bei der Binärzahl 10111, die dezimal der Zahl 23 entspricht, müssen, um die Dezimalzahl 24, binär 11000, zu erhalten, vier Stellen verändert werden. Der Hamming-Abstand entspricht der Anzahl an Stellen einer Kodierung, die verändert werden müssen, um eine benachbarte Zahl zu erhalten [83]. Im Beispiel ist dies vier. Ein Hamming-Abstand größer als eins wird als Hamming-Klippe bezeichnet. Für einen genetischen Algorithmus bedeutet dies, dass um von einer Zahl zu einer benachbarten zu kommen, mehrere Allele

gleichzeitig geändert werden müssen, was bei einem zufälligen Verfahren wie der Mutation nur schwer möglich ist. Es ist also vor allem bei höheren Generationenanzahlen schwierig, die Lösungsvariante nur leicht zu optimieren. Der Gray-Kode ist ebenfalls ein Binärkode, der durch eine spezielle Kodierung immer einen Hamming-Abstand von eins besitzt. Dadurch werden kleine Optimierungsschritte erleichtert.

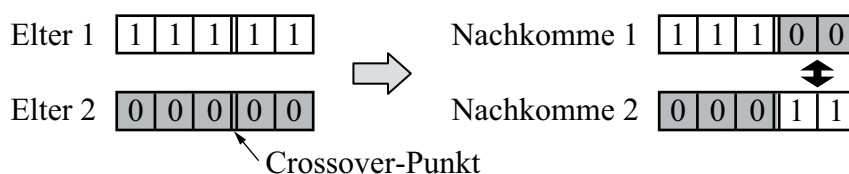
3.1.7 Genetische Operatoren

Die genetischen Operatoren oder Modifikationsoperatoren sind neben der Selektion das aktive Element bei der Optimierung mittels genetischen Algorithmen. Grundsätzlich kann zwischen Rekombination und Mutation unterschieden werden.

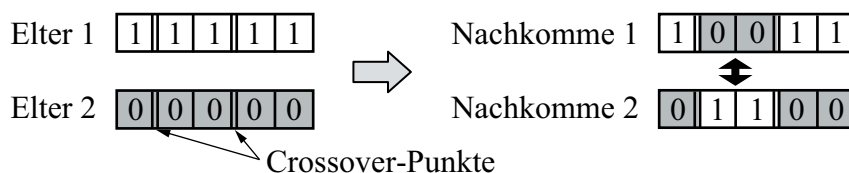
Rekombination

Die Rekombination oder Crossover ist der Hauptmodifikationsoperator bei genetischen Algorithmen. Zur Rekombination werden mindestens zwei Eltern benötigt, die zur Generierung von Nachkommen herangezogen werden. Die Nachkommen werden durch segmentweise Permutation der Eltern erzeugt. Die Rekombination verursacht vor allem in den ersten Generationen eine gute Abdeckung des Lösungsraums, da durch die vielen verschiedenen Lösungsvarianten zu Beginn ständig neue Varianten generiert werden. Mit zunehmender Anzahl an Generationen werden nur noch sehr ähnliche Eltern zur Rekombination verwendet, wodurch wieder sehr ähnliche Nachkommen entstehen. Verschiedene Varianten der Rekombination zeigt Bild 3.2.

1-Punkt-Crossover



N-Punkt-Crossover



Uniform-Crossover

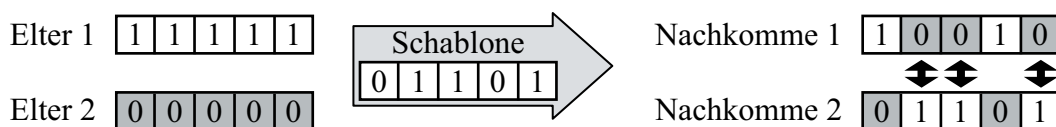


Bild 3.2: Verschiedene Rekombinationsvarianten

Der 1-Punkt-Crossover ist die einfachste Form der Rekombination. Es wird stochastisch ein Crossover-Punkt im Intervall $[1; L_K-1]$ ausgewählt, dabei ist L_K die

Kodelänge. Dieser Crossover-Punkt gilt für jeweils ein Elternpaar. Nach diesem Punkt werden alle Allele der Eltern vertauscht bzw. invertiert, wodurch zwei neue Nachkommen entstehen. Für die Auswahl des Crossover-Punkts ist eine Gleichverteilung anzustreben. Für jedes Elternpaar wird ein neuer Crossover-Punkt bestimmt.

Der N-Punkt-Crossover nutzt $N_C \geq 1$ Crossover-Punkte. Wie beim 1-Punkt-Crossover sind diese stochastisch im Intervall $[1; L_K-1]$ verteilt. Die Anzahl N_C der Crossover-Punkte ist ebenfalls gleichverteilt im Intervall $(1; L_K-1)$. Üblicherweise wird eine gerade Anzahl an Crossover-Punkten erzeugt, die jeweils für ein Elternpaar gelten. Es werden die Allele in allen Abschnitten mit gerader Nummerierung getauscht.

Beim Uniform-Crossover wird zuerst eine Schablone mit der Länge L_K erzeugt. Zur Erzeugung dieser Schablone wird bei jeder Position der Schablone mit einer bestimmten festgelegten Wahrscheinlichkeit p_{cross} überprüft, ob eine Vertauschung stattfinden soll. Ist dies der Fall, befindet sich in der Schablone an dieser Stelle eine „1“. Diese Schablone wird auf beide Eltern angewandt. Es werden immer an der Stelle einer „1“ in der Schablone die Allele getauscht.

In der Literatur [8], [84], [85], [86] sind noch weitere Varianten der Rekombination wie das Shuffle-Crossover und das Diagonal-Crossover beschrieben. Diese Varianten wurden in dieser Arbeit nicht implementiert, weshalb sie hier nur der Vollständigkeit halber erwähnt werden.

Mutation

Der zweite genetische Operator ist die Mutation. Sie ist bei genetischen Algorithmen eher zweitrangig gegenüber der Rekombination. Bei den Evolutionsstrategien und der evolutionären Programmierung ist sie dagegen der einzige Modifikationsoperator. Bei der Mutation wird eine bestimmte Mutationswahrscheinlichkeit p_{mut} festgelegt, die die Wahrscheinlichkeit angibt, mit der ein Allel seinen Wert ändert. Mutationswahrscheinlichkeiten liegen üblicherweise im Bereich von 0,1 % bis 1 %. Von Mühlenbein wurde in [87] vorgeschlagen, die Mutationswahrscheinlichkeit in Abhängigkeit der Kodelänge L_K der Parameter nach Gl. (3.2) zu wählen:

$$p_{mut} = \sqrt{1/L_K} . \quad (3.2)$$

Eine zweite Variante wurde von Jansen in [88] beschrieben. Darin wird das Optimum der Mutationswahrscheinlichkeit nach Gl. (3.3) bestimmt:

$$p_{mut} = \frac{1}{L_K} . \quad (3.3)$$

Mit zunehmender Anzahl an Generationen ähneln sich die Individuen in der Population mehr und mehr. Dadurch verringert sich die Wirksamkeit der Rekombination, um

alle Bereiche des Parameterraums abzubilden. Die Mutation bewirkt im späteren Verlauf eine Erhöhung der Vielfalt an Lösungen und damit eine Verbesserung der Abdeckung des Parameterraums. Eine zu hoch gewählte Mutationswahrscheinlichkeit kann eine Konvergenz der Lösung verhindern, weshalb zwischen der Abdeckung des Parameterraums und der Konvergenz abgewogen werden muss. In Bild 3.3 ist eine Population mit unterschiedlichen Mutationswahrscheinlichkeiten dargestellt. Gut zu erkennen ist die größere Lösungsvielfalt bei höherer Mutationswahrscheinlichkeit.

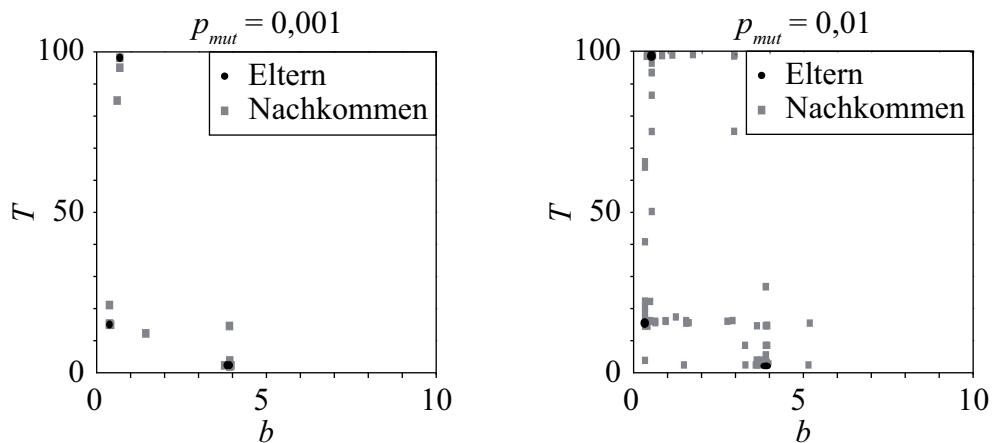


Bild 3.3: Verteilung der Nachkommen bei verschiedenen Mutationswahrscheinlichkeiten und gleicher Populationsgröße

3.1.8 Selektion

Die Selektion hat neben den Modifikationsoperatoren Rekombination und Mutation den größten Einfluss auf die Güte der Lösung und auf die Konvergenz. Die Selektion führt zur Entstehung einer Population höherer Güte. Modifikationen, die rein zufällig stattfinden, bewirken allein keine Verbesserung der Populationsgüte. In Bild 3.1 ist der Ablauf eines genetischen Algorithmus dargestellt mit drei Selektionen pro Zyklus:

- Paarungsselektion
- Populationsselektion
- Umweltselektion

Alle Selektionen dienen einem anderen Zweck, weshalb eine problemspezifische Festlegung der Selektionsvarianten notwendig ist. Bei der Paarungsselektion werden Eltern ausgewählt, die durch Modifikation neue Nachkommen erzeugen sollen. Durch die Populationsselektion entscheidet es sich, ob ausschließlich Nachkommen oder auch Eltern in die nächste Generation übernommen werden. Zudem wird, falls erforderlich, auch das Verhältnis von Eltern und Nachkommen in der Population festgelegt. Zuletzt werden in der Umweltselektion aus der Population die Individuen ausgewählt, welche in der nächsten Generation verwendet werden sollen. Alle Selektionsvarianten werden zu Beginn der Berechnung festgelegt. Für die Paarungs- und Umweltselektion

kommen prinzipiell verschiedene Varianten der Selektion in Frage. Im Folgenden werden die verwendeten Varianten kurz vorgestellt, weitere Varianten finden sich in der einschlägigen Literatur [8], [84], [85], [86].

Stochastische Selektion

Bei der stochastischen Selektion werden zufällig Individuen aus der Population ausgewählt. Dies geschieht als „Ziehen mit Zurücklegen“, wodurch das gleiche Individuum mehrfach ausgewählt werden kann. Bei dieser Variante haben alle Individuen I aus der Population mit einer Menge u_I an Individuen unabhängig von ihrer Güte die gleiche Wahrscheinlichkeit p_S , selektiert zu werden, siehe Gl. (3.4):

$$p_S(I) = \frac{1}{u} \quad \text{für } I = I_j \ (j = 1, \dots, u_I). \quad (3.4)$$

Dadurch kommt es zu einer größeren Vielfalt in der Population, da nicht nur die Individuen mit der höchsten Güte verwendet werden. Nachteilig ist, dass Individuen mit guter Fitness nicht bevorzugt werden, wodurch die Konvergenzgeschwindigkeit sinkt.

Fitnessselektion

Die Fitnessselektion nutzt zur Auswahl der Individuen den Fitnesswert. Dazu werden die Individuen nach ihrer Fitness sortiert und beginnend mit der besten Fitness absteigend selektiert. Der Vorteil dieser Variante ist die sehr hohe Geschwindigkeit, mit der sich die Güte der Population verbessert, da nur die besten Individuen ausgewählt werden. Nachteilig ist, dass die Abdeckung des Lösungsraums schnell eingeschränkt wird. Es werden nur die Besten und damit recht ähnliche Individuen ausgewählt, wodurch die Rekombination an Wirkung verliert.

Fitnessproportionale Selektion

Für die fitnessproportionale Selektion ist ebenfalls der Fitnesswert der Individuen ausschlaggebend. Im Gegensatz zur Fitnessselektion wird eine Wahrscheinlichkeit p_S , dass ein Individuum I_j ausgewählt wird, definiert, die sich linearproportional nach seiner Fitness $Z(I_j)$ richtet, siehe Gl. (3.5):

$$p_S(I_j) = \frac{Z(I_j)}{\sum_{k=1}^u Z(I_k)}. \quad (3.5)$$

Durch die Definition einer Selektionswahrscheinlichkeit p_S für jedes Individuum können auch schlechtere Individuen selektiert werden. Die Güte der Population verbessert sich auch bei dieser Variante sehr schnell. Der Nachteil ist auch hier, wie bei der Fitnessselektion, eine schnelle Verringerung der Lösungsvielfalt und damit der Abdeckung des Lösungsraums, wobei noch eine Wahrscheinlichkeit besteht, auch

schlechtere Individuen zu selektieren. Auch eine Verschlechterung der Fitness der Gesamtpopulation ist möglich.

Rangbasierte Selektion

Bei der rangbasierten Selektion werden die Individuen nach ihrer Fitness in absteigender Reihenfolge geordnet. Daraus ergeben sich die Ränge $r(I)$. Dem Individuum mit der besten Fitness wird der Erwartungswert E_{max} zugewiesen, dem schlechtesten Individuum der Wert E_{min} . Daraus ergibt sich mit Gl. (3.6) die Selektionswahrscheinlichkeit p_s :

$$p_s(I_j) = \frac{1}{u} \left(E_{max} - (E_{max} - E_{min}) \frac{r(I_j) - 1}{u - 1} \right). \quad (3.6)$$

Wegen der Summe der Wahrscheinlichkeiten von $\sum p_s(I_j) = 1$ ergeben sich die Bedingungen $E_{min} = 2 - E_{max}$ und $1 \leq E_{max} \leq 2$. Durch die Wahl von E_{max} kann festgelegt werden, wie stark sich eine bessere Fitness auf die Selektionswahrscheinlichkeit auswirkt. Auch hier ist eine Verschlechterung der Fitness der Gesamtpopulation möglich.

3.1.9 Konvergenzverhalten des genetischen Algorithmus

Die Arbeitsweise des genetischen Algorithmus ist durch den iterativen Prozess aus Selektionen und Anwendung der genetischen Operatoren geprägt. Aus zufällig erzeugten Individuen der Anfangspopulation werden durch den Prozess des genetischen Algorithmus neue Populationen erzeugt, deren Individuen ebenfalls zufällig im Lösungsraum verteilt sind. Die Konvergenz von genetischen Algorithmen ist bei einer endlichen Anzahl an Generationen auf Grund der stochastisch ablaufenden Lösungsfindung nicht garantiert [84]. Zur Bestimmung des Konvergenzverhaltens bei unendlich vielen Generationen werden verschiedene Ansätze, beruhend auf Markov-Ketten, verwendet [89], [90], [91].

Für den kanonischen genetischen Algorithmus nach Holland [64] mit stochastischer Selektion konnte Rudolph in [92] zeigen, dass auch für unendlich viele Generationen keine Konvergenz auf das globale Optimum vorhanden ist. Beim Übergang auf einen genetischen Algorithmus mit einer auf der Güte der Individuen basierenden Selektion verbessert sich das Konvergenzverhalten. Wird in jede neue Population das beste Individuum der vorherigen Generation übernommen, konnte von Rudolph in [92] und Eiben, Aarts und van Hee in [93] eine Konvergenz gegen das globale Optimum bewiesen werden. Diese Art der Selektion wird als Elitismus bezeichnet. In der Fitnessselektion ist der Elitismus durch die direkte Selektion der Individuen nach ihrer Güte umgesetzt.

3.2 Grundlagen künstlicher neuronaler Netze

Künstliche Neuronale Netze (KNN) sind, wie in Kapitel 2.3.4 dargestellt, den Metaheuristiken zuzuordnen. Sie sind eine Methode, die die Funktionsweise eines Gehirns nachahmt. In einem Gehirn arbeiten sehr viele stark miteinander vernetzte Nervenzellen (Neuronen) parallel, wodurch eine sehr effiziente und schnelle Problembearbeitung möglich ist. Im Gegensatz zur Komplexität des Netzwerks sind die einzelnen Nervenzellen sehr einfach aufgebaut. Sie bestehen im Wesentlichen aus Dendriten, dem Nervenzellkörper und dem Axon. In Bild 3.4 ist in eine vereinfachte Darstellung einer Nervenzelle gezeigt.

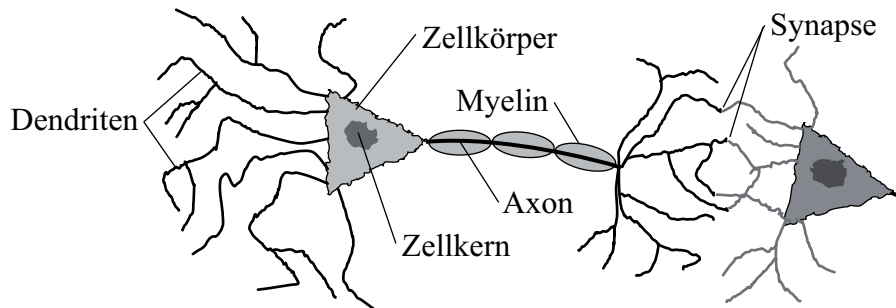


Bild 3.4: Vereinfachte Darstellung einer Nervenzelle

Die stark verzweigten Dendriten dienen der Aufnahme von Informationen vorhergehender Nervenzellen, die dann im Zellkörper verarbeitet werden. Bei Überschreitung eines Schwellwertes wird ein elektrischer Impuls ausgelöst, der sich auf das Axon auswirkt, das als Ausgang dient. Das Axon ist durch das umhüllende Myelin elektrisch isoliert und seinerseits über Synapsen wieder mit Dendriten anderer Nervenzellen verbunden. Dieser prinzipielle Aufbau wird von künstlichen neuronalen Netzen nachgebildet. Es werden viele einfache Recheneinheiten in einer Netzstruktur angeordnet und über Kanten miteinander verbunden. Die Recheneinheiten werden in Anlehnung an das biologische Vorbild Neuronen genannt.

Künstliche neuronale Netze wurden erstmals von McCulloch und Pitts in [94] beschrieben. Daraus entstanden die sogenannten McCulloch-Pitts-Neuronen als Grundelement der neuronalen Netze. In [95] zeigte Hebb die Grundlagen zu Lernvorgängen in Netzstrukturen auf, die als Hebb'sche Lernregel die Grundlage des maschinellen Lernens bildet. Eine zweite wichtige Lernregel ist die von Widrow und Hoff entwickelte Delta-Regel für vorwärtsgerichtete Netze [96]. Eine entscheidende Verbesserung des maschinellen Lernens konnte durch das von Rumelhart, Hinton und Williams in [97] beschriebene Lernverfahren der Backpropagation erreicht werden. In den erwähnten Arbeiten sind die Grundlagen zu den heute verwendeten neuronalen Netzen beschrieben. Eine Reihe weiterer grundlegender Arbeiten und die Historie der Methode sind in der einschlägigen Literatur [98], [99], [100], [101], [102] zu finden.

Auf Grund ihrer vom Gehirn abgeleiteten Struktur werden künstlichen neuronalen Netzen verschiedene Fähigkeiten des biologischen Vorbilds zugeschrieben [103].

- **Lernfähigkeit:**
Künstliche neuronale Netze werden über verschiedene Lernverfahren auf das jeweilige Anwendungsfeld adaptiert. Dazu werden Informationen in der Netzstruktur gespeichert. So können einmal adaptierte Problemstellungen gelöst werden.
- **Verallgemeinerungsfähigkeit:**
Neuronalen Netzen ist es möglich, durch bereits bekannte Eingaben zu unbekanntem Eingaben Lösungen zu generieren. Es können somit für spezielle Eingaben durch die Verallgemeinerung der in der Netzstruktur hinterlegten generellen Eingaben Lösungen gefunden werden.
- **Assoziationsfähigkeit:**
Neuronale Netze besitzen durch ihren Aufbau eine gewisse Fehlertoleranz. Dadurch ist es möglich, auch bei verrauschten Eingangsdaten sinnvolle Ergebnisse zu erhalten.
- **Abstraktionsfähigkeit:**
Aus bereits erlernten Mustern ist es neuronalen Netzen möglich, auch unbekannte Muster zu erkennen und zu klassifizieren.

3.2.1 Neuronale Netze in der Zuverlässigkeitstechnik

Die Anwendung von neuronalen Netzen in der Zuverlässigkeitstechnik nimmt, wie auch die Anwendung von genetischen Algorithmen, zu. Im Folgenden sollen einige Beispiele der Anwendung von neuronalen Netzen gegeben werden.

Die grundlegende Anwendbarkeit von neuronalen Netzen in der Zuverlässigkeitstechnik untersuchten Fink et al. in [104], dabei konnte als Hauptanwendungsbereich die Zuverlässigkeitsprognose festgestellt werden. Maisch setzte in [105] neuronale Netze zur Prognose der Lebensdauer von Getriebeölen ein. Hu, Xie und Ng nutzten in [106] und Karunanithi et al. in [107] neuronale Netze zur Vorhersage der Zuverlässigkeit von Software. Coit und Smith nutzten einen kombinierten Ansatz aus neuronalen Netzen und genetischen Algorithmen zur Bestimmung des kostengünstigsten Redundanzgrades in einem Parallelsystem bei einer vorgegebenen Zuverlässigkeit [108]. Die Bestimmung der Zuverlässigkeit von Telekommunikationsnetzen mit neuronalen Netzen wurde von Altiparmak et al. in [109] untersucht. In gleicher Weise versuchten Sinha und Panday in [110] die Zuverlässigkeit von Öl- und Gaspipelines zu bestimmen.

Van Mahnen erweiterte die Arbeiten von Liu et al. [111] und nutzte neuronale Netze zur Schätzung von Verteilungsparametern von zweiparametrischen Weibullverteilungen

[103]. Dazu untersuchte er verschiedene Felddatensätze und nutzte die Maximum-Likelihood Methode als Vergleich. Es konnten vor allem bei kleinen Stichprobenumfängen bessere Schätzungen über neuronale Netze erreicht werden, als mit der Maximum-Likelihood Methode.

Aufbauend auf den grundlegenden Erfahrungen von Mahrens kann die Parameterschätzung von Einzelverteilungen auf die Analyse von Mischverteilungen und booleschen Systemen erweitert werden. Die guten Ergebnisse von Mahrens für Einzelverteilungen deuten auf eine gute Anwendbarkeit bezüglich des komplexeren Problems mit mehreren Unterverteilungen hin.

3.2.2 Aufbau eines neuronalen Netzwerkes

Neuronale Netze sind gerichtete Graphen, die aus den Elementen Kante und Knoten aufgebaut sind [103]. Dabei ist die Menge der Knoten $N_K = \{n_{11}, \dots, n_{ij}\}$ und die Menge der Kanten $C = \{c_{11}, \dots, c_{ij}\}$ mit $C \subset N_K \times N_K$. Der Index ij bezeichnet die i -te Schicht und den j -ten Knoten oder die j -te Kante. Die Knoten werden im Allgemeinen als Neuronen und die Kanten als Verbindungen bezeichnet.

Ein neuronales Netz ist aus mehreren Teilen aufgebaut, die als Schichten bezeichnet werden. Die Eingabeschicht nimmt die Daten des Eingangsvektors auf. Eine oder mehrere versteckte Schichten oder Zwischenschichten bilden den Übergang zur Ausgabeschicht, die den Ausgabevektor erzeugt. Die einzelnen Schichten bestehen aus Neuronen, die über die Verbindungen miteinander verknüpft sind, siehe Bild 3.5. Die Anzahl an Schichten und Neuronen pro Schicht kann frei gewählt werden.

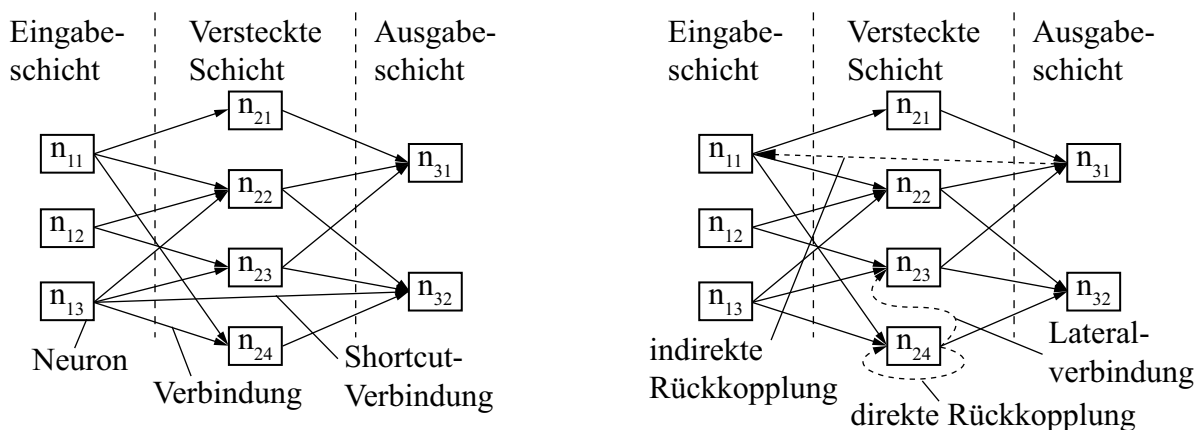


Bild 3.5: Beispielhafte neuronale Netze, ohne Rückkopplung (links), mit Rückkopplung (rechts)

Die Anzahl der Neuronen in der Eingabeschicht ist durch die Größe des Eingangsvektors problemspezifisch festgelegt. Die Anzahl der Neuronen in der Ausgabeschicht ist abhängig von der Größe des Ausgangsvektors. Die Anzahl der Neuronen in der Zwi-

schenschicht und die Anzahl der Zwischenschichten muss dem jeweiligen Problem angepasst werden.

Es kann zwischen Netzen ohne Rückkopplung (Feedforward-Netze), siehe Bild 3.5 links, und Netzen mit Rückkopplung (Feedback-Netze), siehe Bild 3.5 rechts, unterschieden werden. Sind in einem Netz ohne Rückkopplung nur Verbindungen zwischen zwei benachbarten Schichten vorhanden, ist dies ein Netz 1. Ordnung. Sind Verbindungen vorhanden, die eine oder mehrere Schichten überspringen, sogenannte Shortcut-Verbindungen (Bild 3.5 links), ist es ein Netz 2. Ordnung. Bei Netzen mit Rückkopplung sind Verbindungen zwischen Neuronen höherer Schichten mit Neuronen niedriger Schichten oder derselben Schicht vorhanden. Hier werden direkte Rückkopplungen, indirekte Rückkopplungen und Lateralverbindungen unterschieden (Bild 3.5 rechts). Alle Verbindungen haben eine durch den Pfeil angedeutete Übertragungsrichtung. Um die Stärke der Verbindung zwischen zwei Neuronen anzugeben, sind die Verbindungen mit einer Gewichtung w versehen, weshalb neuronale Netze auch als gewichtete Graphen bezeichnet werden [103]. Die Veränderung der Gewichte bewirkt eine starke Veränderung des Zusammenhangs zwischen Eingangs- und Ausgabevektor.

3.2.3 Das Neuron

Die zentralen Elemente eines neuronalen Netzes sind die Neuronen. In Bild 3.6 ist der Aufbau eines Neurons dargestellt.

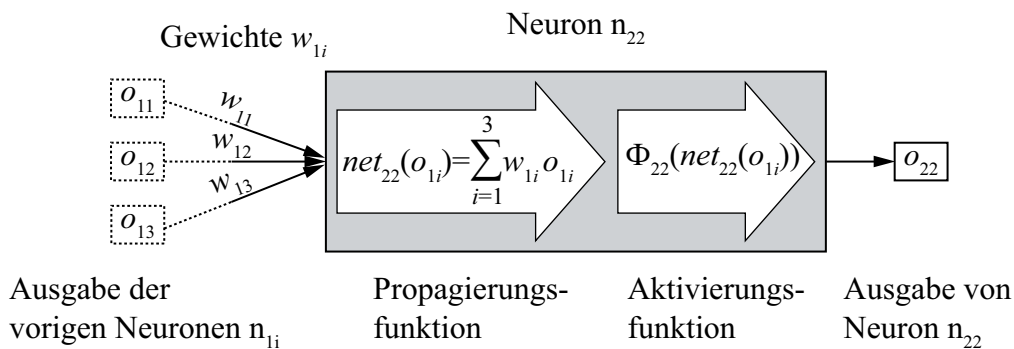


Bild 3.6: Aufbau und Funktionsweise eines Neurons

Als Eingangsgrößen eines Neurons dienen die Werte der gewichteten Verbindungen. Die einzelnen Werte werden der Propagierungsfunktion $net_j(o_i)$ zugeführt. Die Propagierungsfunktion kann eine Summation oder Multiplikation der gewichteten Eingangswerte o_i des Neurons sein. Das Ergebnis der Propagierungsfunktion dient als Eingangsgröße für die Aktivierungsfunktion $\Phi_j(net_j)$.

$$o_j = \Phi_j(net_j) = \Phi_j\left(\sum_{i=1}^k w_{ij} o_i\right) \quad (3.7)$$

Die Aktivierungsfunktion beschreibt die Aktivität des Neurons, die bei der Ausgabe von binären Werten ab einem Schwellwert S einen Ausgangswert erzeugt.

$$\Phi_j(\text{net}_j) = \begin{cases} 0, & \text{net}_j < S \\ 1, & \text{net}_j > S \end{cases} \quad (3.8)$$

Für Ausgaben von reellen Zahlenwerten werden verschiedene Aktivierungsfunktionen verwendet, die zur Anwendung bestimmter Trainingsverfahren differenzierbar sein müssen. Bei nichtlinearen Funktionen werden nichtlineare Aktivierungsfunktionen verwendet. Häufig kommt hierbei die Sigmoidfunktion zum Einsatz, siehe Gl. (3.9):

$$\Phi_j(\text{net}_j) = \frac{1}{1 + e^{-\alpha \cdot \text{net}_j}}. \quad (3.9)$$

Sigmoidfunktionen sind nichtnegativ, streng monoton steigend und streben asymptotisch gegen Null für $\text{net}_j(o_i) \rightarrow -\infty$ und gegen Eins für $\text{net}_j(o_i) \rightarrow \infty$ [112].

3.2.4 Netztraining

Das Netztraining bezeichnet die Anpassung eines Netzes auf eine bestimmte Anwendung. Zur Anpassung des Netzes gibt es verschiedene Herangehensweisen:

- Ergänzen oder Entfernen von Verbindungen oder Neuronen,
- Variieren der Verbindungsgewichte,
- Variieren der internen Parameter des Neurons, z. B. der Aktivierungsfunktion.

Die am häufigsten umgesetzte Variante ist die Variation der Verbindungsgewichte [98]. Mittels des Netztrainings wird versucht, die Verbindungsgewichte so anzupassen, dass ein gefordertes Ergebnis erreicht wird. Beim sogenannten „überwachten Lernen“ sind Ein- und Ausgangsdaten eines Datensatzes bekannt. Dieser Datensatz wird Trainingsdatensatz genannt. Es existieren neben dem überwachten Lernen noch die Varianten des nicht überwachten Lernens, bei denen die Ausgangsdaten nicht bekannt sind [100]. Das nicht überwachte Lernen wird in dieser Arbeit nicht weiter betrachtet.

Die Grundlage des überwachten Lernens stellt die Hebb'sche Lernregel dar. Nach Hebb ist die Verbindung zwischen zwei Neuronen umso stärker, je öfter sie gemeinsam aktiv sind. Die Gewichtung der Verbindung w_{ij} zwischen Neuron i und j muss dann nach Gl (3.10) um Δw_{ij} verstärkt werden:

$$\Delta w_{ij} = \eta \Phi_j o_i. \quad (3.10)$$

Hierbei ist η die konstante Lernrate, die die Höhe der Veränderung des Verbindungsgewichts Δw_{ij} für einen Lernschritt definiert. Die Wahl der Lernrate hat einen direkten Einfluss auf die Konvergenz des Verfahrens. Eine zu große Lernrate kann zu einem

Überspringen von Optima führen und eine Divergenz des Verfahrens verursachen, eine geringe Lernrate verringert die Konvergenzgeschwindigkeit des Verfahrens.

Die zweite Grundlage des überwachten Lernens ist die Delta-Regel für Netze ohne Rückkopplung. Dabei wird die Verbindungsgewichtsänderung Δw_{ij} auf Basis der Abweichung der Ist-Ausgabe o_j von der Soll-Ausgabe \bar{o}_j für das Neuron j bestimmt.

$$\Delta w_{ij} = \eta o_i (\bar{o}_j - o_j) \quad (3.11)$$

In Gl. (3.11) ist o_i die Ausgabe des Vorgängerneurons.

3.2.5 Lernverfahren

Auf Basis der Hebb'schen Lernregel und der Delta-Regel entstanden verschiedene Lernverfahren wie Backpropagation [97], Quickprop [113] und Resilient Propagation [114], [115]. Das am häufigsten verwendete Lernverfahren ist der Backpropagation-Algorithmus. In dieser Arbeit wird im Folgenden nur das Backpropagation-Verfahren betrachtet.

Das überwachte Lernen oder Netztraining stellt eine Optimierung der Kantengewichte w_{ij} dar. Um eine Optimierung durchführen zu können, muss eine Zielfunktion definiert werden. Im Bereich der neuronalen Netze wird diese Funktion Energiefunktion Q genannt. Die Energiefunktion ist die Summe aller Abweichungen der Ist-Ausgaben $o_{a,j}$ von den Soll-Ausgaben $\bar{o}_{a,j}$ aller Neuronen in der Ausgangserschicht j für alle Trainingsdatensätze a .

$$Q = \sum_a Q_a = \frac{1}{2} \sum_a \sum_j (\bar{o}_{a,j} - o_{a,j})^2 \quad (3.12)$$

Das Ziel des Netztrainings ist es, die Energiefunktion durch die Anpassung der Verbindungsgewichte w_{ij} zu minimieren. Dies führt zu der Vorgehensweise eines Gradientenabstiegsverfahrens, bei dem w_{ij} schrittweise um Δw_{ij} verändert wird.

$$\Delta_a w_{ij} = - \frac{\partial Q_a}{\partial w_{ij}} \quad (3.13)$$

Die Anwendung der Kettenregel führt zu den partiellen Ableitungen

$$\frac{\partial Q_a}{\partial w_{ij}} = \frac{\partial Q_a}{\partial net_{a,j}} \cdot \frac{\partial net_{a,j}}{\partial w_{ij}}. \quad (3.14)$$

In Gl. (3.14) ist $net_{a,j}$ die Propagierungsfunktion von Neuron j beim Trainingsdatensatz a . Der erste Faktor beschreibt die Veränderung von Q_a in Abhängigkeit der Ausgabe von Neuron j . Der zweite Faktor ist die Ist-Ausgabe von Neuron j beim Trainingsdatensatz a , siehe Gl. (3.15).

$$\frac{\partial net_{a,j}}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \sum_k w_{jk} o_{a,k} = o_{a,j} \quad (3.15)$$

Die Ist-Ausgabe $o_{a,j}$ ergibt sich aus der Summe der Eingänge $w_{jk} o_{a,k}$. Der erste Faktor von Gl. (3.14) wird nach Gl. (3.16) definiert als

$$\delta_{a,j} = -\frac{\partial Q_a}{\partial net_{a,j}}. \quad (3.16)$$

Damit ergibt sich mit Gl. (3.15) und Gl. (3.16) in Gl. (3.14)

$$-\frac{\partial Q_a}{\partial w_{ij}} = \delta_{a,j} o_{a,j}. \quad (3.17)$$

Zur Umsetzung des Gradientenabstiegs von Q ergibt sich die Gewichtsänderung unter Einbeziehung der Lernrate η zu

$$\Delta_a w_{ij} = \eta \delta_{a,j} o_{a,j}. \quad (3.18)$$

Die Ableitung von $\delta_{a,j}$ und die Anwendung der Kettenregel führt zu unterschiedlichen δ für Ausgabeneuronen und für Neuronen der Zwischenschicht.

$$\delta_{a,j} = \begin{cases} \Phi'_j(net_{a,j})(\bar{o}_{a,j} - o_{a,j}) & \text{für Neuron } j \in \text{Ausgabeschicht} \\ \Phi'_j(net_{a,j}) \sum_k \delta_{ak} w_{jk} & \text{für Neuron } j \in \text{Zwischenschicht} \end{cases} \quad (3.19)$$

Mit den vorgestellten Gleichungen kann das überwachte Lernen durchgeführt werden.

Der Ablauf des Netztrainings gliedert sich in drei Phasen:

- Der Eingabevektor wird in die Eingabeschicht eingespeist.
- Der Fehler für den Eingabevektor wird berechnet.
- Die Gewichtungen werden über das Lernverfahren, beginnend mit der Ausgabeschicht, Schicht für Schicht angepasst.

Beim Netztraining kann es durch eine zu große Anzahl an Trainingsdaten zu einer Überanpassung (overfitting) des Netzes kommen, wodurch die Trainingsdaten sehr gut wiedergegeben werden, die Evaluierungsdaten jedoch schlechte Lösungen liefern. Um diesem Sachverhalt zu begegnen, wurde von Rumelhart et al. in [97] der Momentum- oder Trägheitsterm eingeführt.

$$\Delta_a w_{ij}(t+1) = \eta \delta_{a,j} o_{a,j} + \mu \Delta_a w_{ij}(t) \quad (3.20)$$

Der Parameter μ ist eine frei wählbare Konstante im Intervall $[0, 1]$. Damit wird der Gradient im vorherigen Schritt t bei der Berechnung der Gewichtung des aktuellen Schrittes $t+1$ berücksichtigt, wodurch das Konvergieren zu lokalen Minima

erschwert wird. Zusätzlich kann mit einer kontinuierlichen Verringerung der Lernrate mit zunehmender Schrittzahl t_i einer Überanpassung entgegen gewirkt werden [112].

3.2.6 Vorgehensweise bei der Parameterschätzung von realen Daten

Zur Schätzung von Parametern realer Ausfalldaten müssen künstliche neuronale Netze für jeden Datensatz neu erstellt werden. Der Ablauf der Parameterschätzung lässt sich in verschiedene Schritte untergliedern, die in Bild 3.7 dargestellt sind.

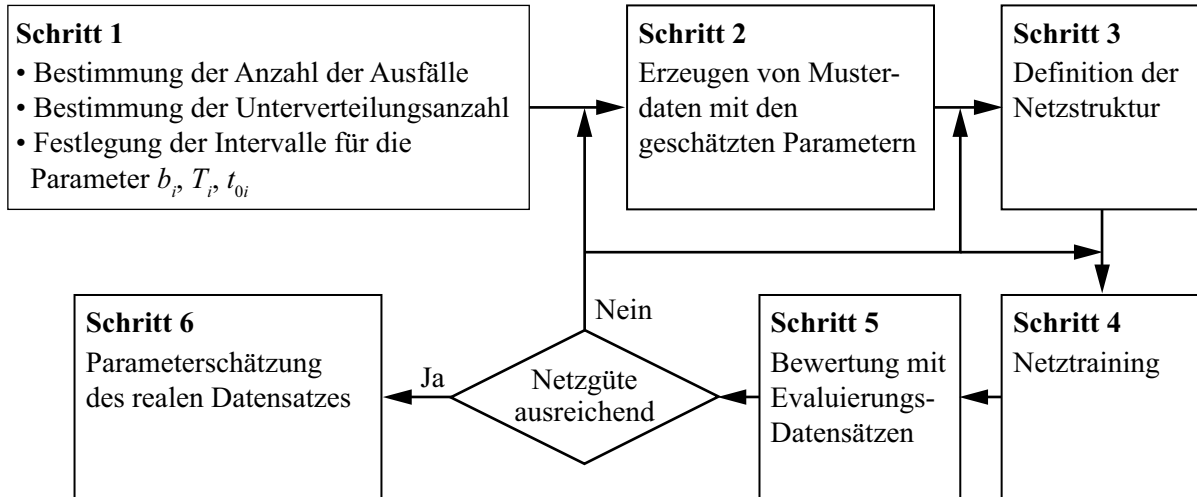


Bild 3.7: Ablauf bei der Parameterschätzung mittels künstlicher neuronaler Netze

Der erste Schritt beinhaltet die Eingrenzung des Parameterraums für die Erstellung der Trainings- und Evaluierungsdaten. Aus den Ausfalldaten des realen Datensatzes müssen die verschiedenen Parameter zur Generierung der Musterdaten ermittelt werden. Die Anzahl an Ausfällen des Datensatzes definiert die Anzahl an Neuronen in der Eingabeschicht. Die Anzahl und der Typ der Unterverteilungen oder Ausfallmechanismen müssen für die Anzahl an Neuronen in der Ausgabeschicht bestimmt werden. Mit diesen Randbedingungen müssen Musterverteilungen und daraus Musterdatensätze erzeugt werden, die den erwarteten Bereich der realen Daten einschließen. Dazu müssen die Werte für die charakteristischen Lebensdauern T_i , die eventuell vorhandenen ausfallfreien Zeiten t_{0i} und die Bereiche der Formparameter b_i geschätzt werden. Mit diesen Werten werden in Schritt zwei Musterdaten erzeugt, die möglichst viele potentielle Kombinationen von Parametern beinhalten.

In Schritt drei wird, basierend auf den Werten aus Schritt eins, eine Netzstruktur mit einer festgelegten Anzahl an Neuronen in der Eingabe- und Ausgabeschicht definiert. Zusätzlich muss die Anzahl der versteckten Schichten und der Neuronen pro Schicht festgelegt werden. Ferner muss noch definiert werden, ob ein Netz erster oder zweiter Ordnung verwendet werden soll. Zuletzt wird die Art der Propagierungs- und Aktivierungsfunktion bestimmt.

Zum Training und zur Evaluierung des Netzes werden die erzeugten Musterdatensätze in zwei Gruppen aufgeteilt. Mit einem Teil der Musterdaten, den Trainingsdaten, wird in Schritt vier das Netz mittels eines Verfahrens des überwachten Lernens, beispielsweise Backpropagation, trainiert. Im anschließenden fünften Schritt wird mit dem zweiten Teil der Musterdaten, den Evaluierungsdaten, die Güte des Netzes und der Ergebnisse bewertet. Bei nicht ausreichender Güte der Ergebnisse können iterativ die Schritte vier, drei und zwei angepasst werden. Sinnvoll ist es, zuerst in Schritt vier das Netztraining zu verändern und sukzessive bei weiteren Optimierungsläufen die Parameter in Schritt drei und zwei anzupassen.

In Schritt sechs können die Parameter der Verteilungen der realen Daten mittels des künstlichen neuronalen Netzes geschätzt werden. Die Güte dieser Schätzung hängt von den vorherigen Schritten ab, wobei nicht generell ein globales Optimum der Parameter erreicht werden kann.

3.3 Vergleich der Anwendung von genetischen Algorithmen und neuronalen Netzen

Genetische Algorithmen und künstliche neuronale Netze sind von ihrer Methodik her prinzipiell geeignet, die in dieser Arbeit betrachtete Problemstellung zu lösen. Bei beiden Methoden ist die Vielzahl der zu schätzenden Parameter von untergeordneter Bedeutung.

Bei genetischen Algorithmen ist der kritische Punkt zur Parameterschätzung von gemischten Verteilungen die Definition der Fitnessfunktion. Sie muss so gewählt werden, dass ein Vergleich zwischen den verschiedenen Lösungsvarianten erfolgen kann. Zusätzlich kommt der Wahl der Kodierung eine wichtige Rolle zu. Sie muss so getroffen werden, dass alle Bereiche des Lösungsraums abgedeckt werden können. Die Wirksamkeit der genetischen Operatoren ist von der Kodierung der Parameter abhängig. Alle weiteren Variationsmöglichkeiten des genetischen Algorithmus, wie die Wahl der Rekombinationsvariante, der Mutationswahrscheinlichkeit und der Selektionsmethoden sind problemspezifisch festzulegen. Eine schlechte Wahl der genetischen Operatoren und Selektionsmethoden beeinträchtigt die Konvergenz und die Güte der Lösung, es wird jedoch in jedem Fall eine Lösung generiert. Eine Anpassung des genetischen Algorithmus zur Parameterschätzung von verschiedenen gemischten Verteilungen ist über die Fitnessfunktion möglich. Zudem besteht die Möglichkeit, zensierte Daten über die in Abschnitt 2.1 vorgestellten Methoden mit einzubeziehen.

Nachteilig bei genetischen Algorithmen ist die im Vergleich zu analytischen Verfahren, wie dem Newton-Raphson-Verfahren, geringe Konvergenzgeschwindigkeit. In Abhängigkeit von der Generationenanzahl und der Populationsgröße können lange

Rechenzeiten auftreten. Zusätzlich ist das Finden eines globalen Optimums innerhalb einer endlichen Zeit nicht garantiert, nach [75] kann eine „akzeptable Lösung“ in einer akzeptablen Zeit erreicht werden.

Bei künstlichen neuronalen Netzen ist der kritische Punkt die Generierung geeigneter Trainingsdaten und das Netztraining. Die Trainingsdaten müssen den Bereich, in dem die vermuteten Parameter der Verteilungen liegen, möglichst vollständig abdecken. Das Netztraining mit Hilfe der Trainingsdaten muss so durchgeführt werden, dass es zu keiner Überanpassung des Netzes kommt. Es muss ein Kompromiss bei der Wahl der Trainingsdaten gefunden werden, bei dem der Parameterraum möglichst vollständig abdeckt wird, aber aufgrund der großen Anzahl an Trainingsdaten keine Überanpassung auftritt. Ist ein Netz für eine Problemstellung generiert und trainiert, lassen sich damit sehr schnell Lösungen des Problems finden.

Nachteilig bei neuronalen Netzen ist der starre Aufbau des Netzes. Es muss für jede Anzahl an Eingangsparametern, in diesem Fall die Ausfallzeiten, ein eigenes Netz mit der entsprechenden Anzahl an Eingangsneuronen generiert werden. Selbiges gilt für die Anzahl an Ausgangsneuronen, die in Abhängigkeit der Verteilungsfunktion und der Anzahl an Unterverteilungen festgelegt werden muss. Für jedes Netz müssen wieder geeignete Trainingsdaten generiert und das Netz mit diesen trainiert werden. Der Problematik mit ungleichen Stichprobengrößen begegnen Liu et al. [111] mit einer Klassierung der Eingangsdaten, was jedoch zu einem Informationsverlust bei der Parameterschätzung führt. Wie bei genetischen Algorithmen kann bei der gefundenen Lösung nicht das globale Optimum erwartet werden.

Im Vergleich der beiden Methoden bieten die genetischen Algorithmen die universellere Anwendbarkeit. Es ist keine Abhängigkeit von der Anzahl der Eingangsparameter, also der Stichprobengröße, vorhanden. Zudem ist die Anzahl der Ausgangsparameter nicht durch den Ablauf des Algorithmus limitiert. Es ist auch möglich, die Anzahl der Unterverteilungen als zusätzlichen zu schätzenden Parameter einzuführen. Diese Vorteile wiegen den Nachteil der höheren Rechenzeit gegenüber einem trainierten neuronalen Netz auf. Der Vorteil der Schnelligkeit der neuronalen Netze ist nur vorhanden, wenn schon ein auf die Problemstellung trainiertes Netz genutzt werden kann.

Durch die Analyse der Funktionsweise und des Aufbaus der genetischen Algorithmen und der künstlichen neuronalen Netze konnte eine prinzipielle Eignung beider Verfahren für die Analyse von gemischten Verteilungen aufgezeigt werden. Parameterschätzungen von booleschen Seriensystemen mit genetischen Algorithmen und künstlichen neuronalen Netzen sind in [116] und in [117] beschrieben. Wegen der universelleren Anwendbarkeit wird im Rahmen dieser Arbeit die Parameterschätzung mittels genetischer Algorithmen genutzt.

4 Parameterschätzung mit genetischen Algorithmen

Eine allgemeine Beschreibung von genetischen Algorithmen und deren generellem Konzept und Ablauf findet sich in Kapitel 3.1. Die Anwendung der genetischen Algorithmen zur Schätzung von Parametern von gemischten Verteilungen ist Inhalt der folgenden Abschnitte. Anhand der in Kapitel 3.1 gezeigten Grundlagen werden die genetischen Algorithmen auf das vorliegende Problem der Schätzung der Verteilungsparameter angepasst.

Zur Schätzung der Parameter müssen verschiedene Vorüberlegungen zum zu bearbeitenden Problem angestellt werden. Dazu zählt die Festlegung der notwendigen Eingangsgrößen für den Algorithmus und die Definition des zu bearbeitenden Problems, siehe Bild 4.1. Nachdem die Vorverarbeitung abgeschlossen ist, läuft der eigentliche genetische Algorithmus im Verarbeitungsschritt ab. Die Ergebnisse des genetischen Algorithmus werden im Nachverarbeitungsschritt zur weiteren Nutzung aufbereitet.

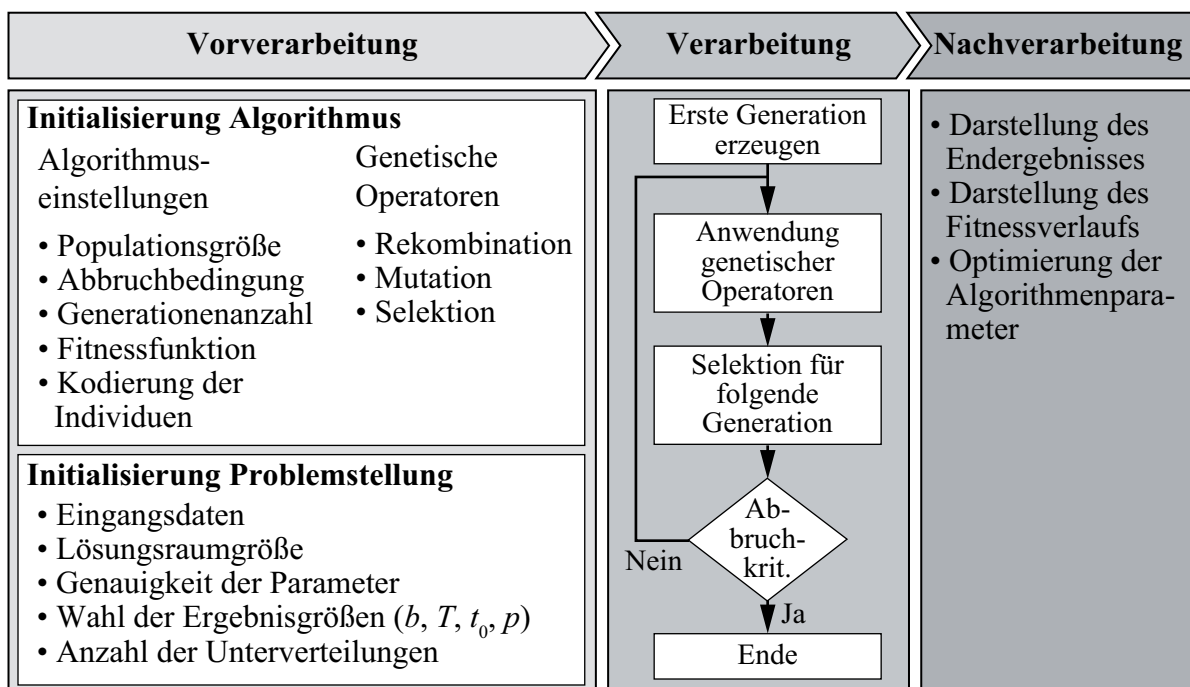


Bild 4.1: Ablauf der Parameterschätzung von gemischten Verteilungen mit genetischen Algorithmen

Die einzelnen Initialisierungsgrößen des Algorithmus und der Problemstellung werden in den nachfolgenden Abschnitten erläutert und deren Bedeutung für die Güte der Parameterschätzung diskutiert. Die Güte der Parameterschätzung wird anhand der in Kapitel 4.1 vorgestellten Datensätze für verschiedene Varianten der betrachteten Größen untersucht.

In Bild 4.2 ist der schematische Ablauf für die Parameterschätzung dargestellt.

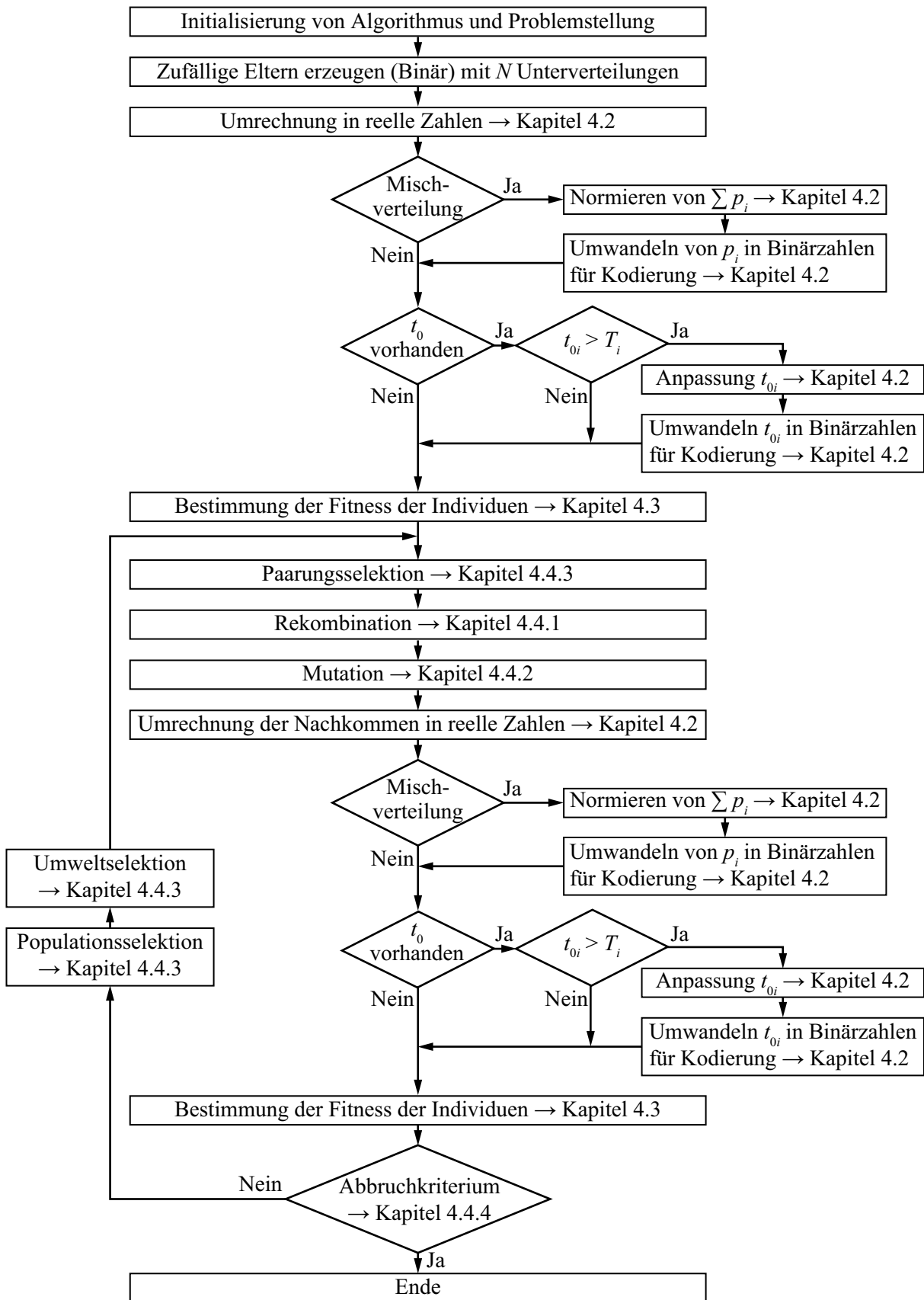


Bild 4.2: Ablauf der Parameterschätzung von gemischten Verteilungen mit genetischen Algorithmen

Der Ablauf des genetischen Algorithmus wurde in der Software MATLAB[®] von Mathworks[®] implementiert. Auf die einzelnen, in Bild 4.2 dargestellten, Ablaufschritte wird in den folgenden Kapiteln Bezug genommen. Hierbei werden auch die erreichbare Güte der Parameterschätzung und die Konvergenz in Abhängigkeit von den untersuchten Möglichkeiten diskutiert. Dazu werden im ersten Schritt Datensätze von Einfach- und Mischverteilungen genutzt.

Zum Vergleich der unterschiedlichen Parameterkombinationen dient in der Nachbereitung der Berechnungen die Abweichung der gefundenen Verteilungskurve von den zu schätzenden Punkten. Dazu wird die Summe D_S der minimalen Abstandsquadrate D_i der gegebenen Datenpunkte zur Verteilungskurve verwendet, siehe Gl. (4.1).

$$D_S = \sum_{i=1}^n D_i^2 \quad (4.1)$$

Zur Berechnung wird wegen der Monotonie das Minimum des Quadrats der Abweichung D_i gesucht. Damit ergibt sich der Abstand D_i allgemein nach Gl. (4.2).

$$D_i^2 = (x_i - x_{pi})^2 + (F(x_i) - y_{pi})^2 \quad (4.2)$$

Darin ist x_i der Abszissenwert des Punktes auf der Kurve, der den geringsten Abstand vom aktuell betrachteten Datenpunkt hat, $F(x_i)$ der Funktionswert der Verteilung an der Stelle x_i und x_{pi} und y_{pi} die Abszissen- und Ordinatenwerte des Datenpunktes. Aus Gl. (4.2) ergibt sich die Ableitung für das Minimum von D_i

$$\frac{dD_i^2}{dx_i} = 2(x_i - x_{pi}) + 2f(x_i)(F(x_i) - y_{pi}). \quad (4.3)$$

Mit Gl. (2.13) und Gl. (2.14) für Mischverteilungen und $x_i = t_i$ ergibt sich

$$\frac{dD_i^2}{dt_i} = 2(t_i - x_{pi}) + 2 \sum_{j=1}^N (f_j(t_i) p_j) (\sum_{j=1}^N (F_j(t_i) p_j) - y_{pi}). \quad (4.4)$$

Der Nulldurchgang wird über die Bisektionsmethode bestimmt. Diese Methode ist bei stückweise stetigen Mischverteilungen mit ausfallfreier Zeit t_0 anwendbar.

4.1 Mischverteilungen zur Evaluation der Parameter des genetischen Algorithmus

Für die Evaluation der Parameter des genetischen Algorithmus werden verschiedene repräsentative Verteilungen verwendet. Sie basieren auf den von Jiang und Kececioglu in [9] vorgestellten Mischverteilungen. Eine Beschreibung der Einteilung der Mischverteilungen findet sich in Kapitel 5.1.1. Im Folgenden sind in Bild 4.3 und Bild 4.4 die verschiedenen Typen der Mischverteilungen dargestellt. Erzeugt werden diese mit

den in Tabelle 4.1 dargestellten Werten, wobei die Parameter b_1 , b_2 , T_1 und p_1 konstant gehalten wurden und nur T_2 variiert wurde.

Tabelle 4.1: Parameter der Mischverteilungen nach Jiang und Kececioglu [9]

	b_1	T_1	p_1	b_2	T_2	p_2
Typ A	1	1	0,3	3	0,01	0,7
Typ B	1	1	0,3	3	0,2	0,7
Typ C	1	1	0,3	3	1	0,7
Typ D	1	1	0,3	3	4	0,7
Typ E	1	1	0,3	3	20	0,7
Typ F	1	1	0,3	3	100	0,7

Durch die Variation von T_2 entstehen verschiedene Formen von Mischverteilungen. Die Typen A bis C sind in Bild 4.3 mit den zugehörigen Unterverteilungen dargestellt.

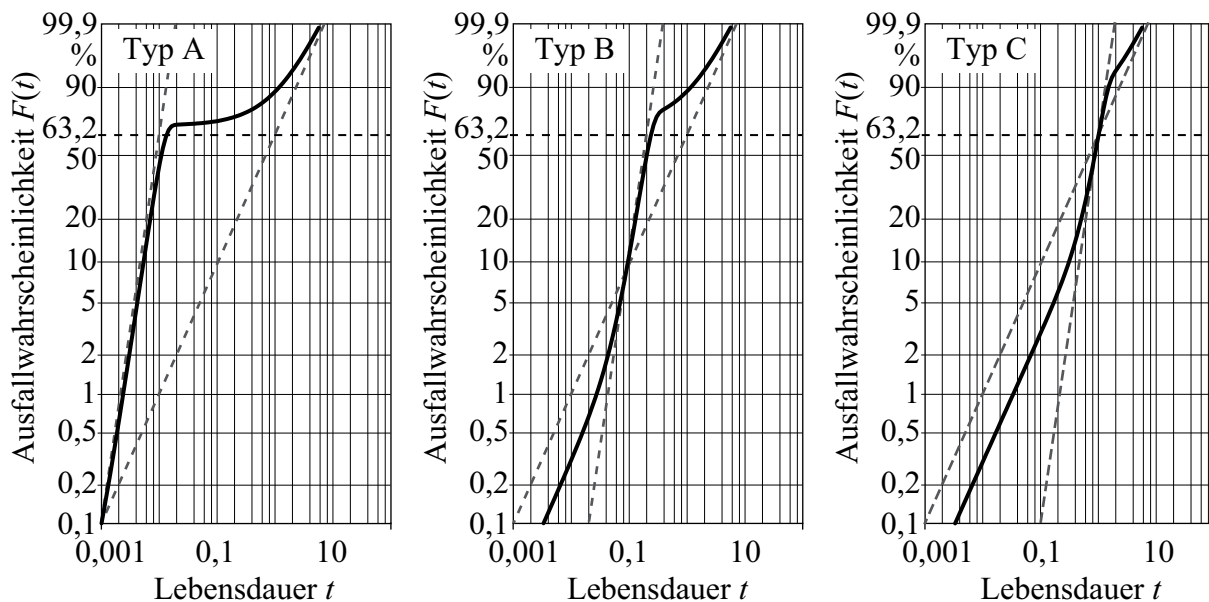


Bild 4.3: Mischverteilungstypen A, B und C nach Jiang und Kececioglu [9]

Die Merkmale der Typen A bis C sind:

- Typ A entsteht, wenn die charakteristische Lebensdauer $T_2 \ll T_1$ ist. Dadurch ist der Überlappungsbereich der Dichtefunktionen der Unterverteilungen gering. Es tritt eine Stufe im Verlauf der Mischverteilung auf.
- Typ B ist ähnlich zu Typ A mit $T_2 < T_1$. Der Überlappungsbereich ist gegenüber Typ A vergrößert. Die Stufe im Verlauf wird zu einer Knickstelle und der Gesamtverlauf ist leicht s-förmig.
- Typ C entsteht, wenn für die charakteristischen Lebensdauern $T_2 \approx T_1$ gilt und die Formparameter sich stark unterscheiden $b_2 > b_1$. Der Überlappungsbereich ist sehr groß. Die Stufe im Verlauf schwächt sich zu einer leichten Knickstelle im s-förmigen Verlauf ab.

Allgemein tritt bei den Typen A bis C die Stufe im Verlauf der Mischverteilung eher bei höheren Ausfallwahrscheinlichkeiten auf. Bei den Typen D bis F verschiebt sich die Stufe, bedingt durch $T_2 > T_1$, zu geringeren Ausfallwahrscheinlichkeiten. Die Mischverteilungstypen D bis F sind mit den zugehörigen Unterverteilungen in Bild 4.4 dargestellt.

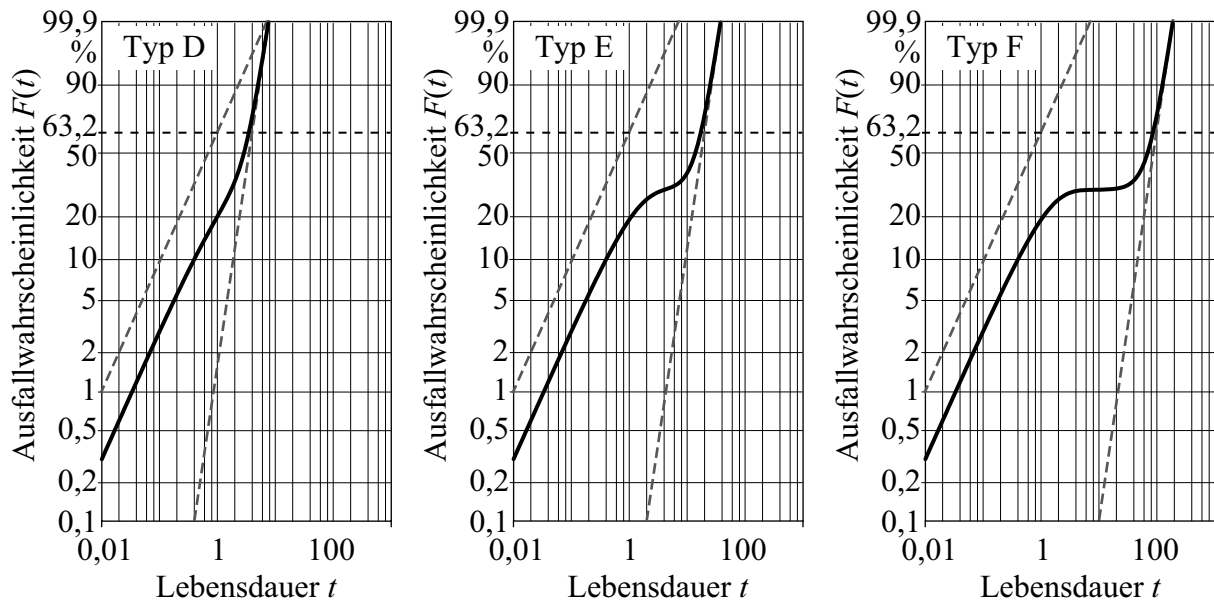


Bild 4.4: Mischverteilungstypen D, E und F nach Jiang und Kececioglu [9]

Die Merkmale der Typen D bis F sind:

- Typ D tritt auf, wenn der Formparameter $b_2 > b_1$ und die charakteristische Lebensdauer T_2 etwas größer ist als T_1 . Es tritt ein Knick im Verlauf auf, der Gesamtverlauf ist nicht mehr s-förmig.
- Typ E ist vorhanden, wenn, ähnlich wie bei Typ D, die charakteristische Lebensdauer $T_2 > T_1$ ist. Es tritt eine Stufe im Verlauf auf.
- Typ F ist ähnlich zu Typ E, wobei hier gilt $T_2 \gg T_1$. Es zeigt sich im Verlauf der Mischverteilung ein Sattelpunkt.

Mit den hier vorgestellten Mischverteilungstypen werden Optimierungsläufe für die Parameter des genetischen Algorithmus durchgeführt. Anhand der Ergebnisse werden geeignete Parameter für die Schätzung von gemischten Verteilungen abgeleitet.

4.2 Definition der Kodierung

Die Grundlage für die Verarbeitung von Daten mit genetischen Algorithmen ist die Definition der Kodierung. Die Kodierung umfasst alle zur Charakterisierung eines Individuums erforderlichen Größen. Dies sind für Mischverteilungen die Parameter der Weibullverteilung b_i , T_i , t_{0i} und die Gewichtungsfaktoren p_i . Für boolesche Systeme entfallen die Gewichtungsparameter p_i . Im einfachsten Fall können die Individuen

reellwertig kodiert werden. Im Folgenden werden jedoch nur die für genetische Algorithmen am häufigsten verwendete Binärkodierung und die Gray-Kodierung eingesetzt.

Die Kodierung der Parameter erfolgt sequentiell. Dabei werden in einer vorgegebenen Reihenfolge die binärkodierten Parameter der einzelnen Verteilungen aneinander angehängt und somit eine Datenreihe erzeugt, siehe Bild 4.5.

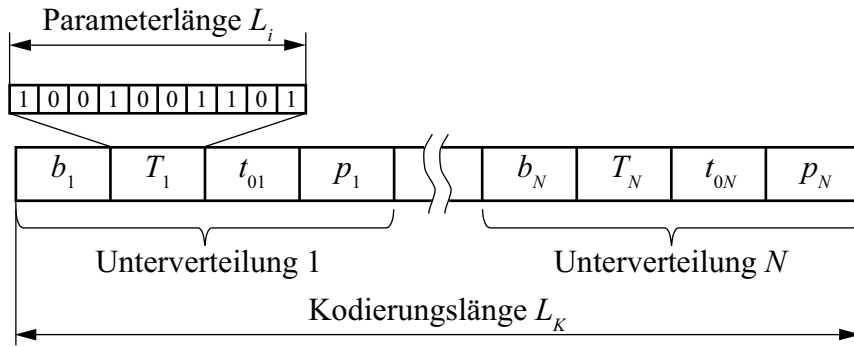


Bild 4.5: Aufbau der binären Kodierung eines Individuums bei Mischverteilungen mit N Unterverteilungen

Die Länge des Codes eines einzelnen Parameters, die Parameterlänge L_i , ergibt sich nach Gl. (4.5) aus der Lösungsraumgröße für den betrachteten Parameter x , also dem Definitionsbereich, und der erforderlichen Genauigkeit m_x des Parameters.

$$L_i = \left\lceil \frac{\ln((x_o - x_u)10^{m_x})}{\ln 2} \right\rceil \quad (4.5)$$

Der Definitionsbereich eines Parameters x ist definiert durch das Intervall $[x_u, x_o]$. Die erforderliche Genauigkeit m_x ist die erforderliche Anzahl an Nachkommastellen der Lösung. Die Aufrundung der Lösung führt immer zu ganzzahligen Werten für L_i . Die Werte für m_x und die Lösungsraumgröße werden bei der Initialisierung der Problemstellung für jeden Parameter definiert. Die Kodierungslänge L_K berechnet sich nach Gl. (4.6) mit der Anzahl der Parameter pro Verteilung n_p .

$$L_K = N \sum_{i=1}^{n_p} L_i \quad (4.6)$$

Die Parameter in der binären Darstellung werden für die genetischen Operatoren, Rekombination und Mutation, genutzt. Die reellwertige Darstellung der Parameter dient zur Weiterverarbeitung und zur Bestimmung der Fitness. Die Berechnung der Parameter in reellen Zahlen aus der binären Kodierung erfolgt mit Gl. (4.7).

$$x_{dec} = x_u + \frac{x_o - x_u}{2^{L_x} - 1} \sum_{i=1}^{L_x} (X_{(L_x-i+1)} 2^{i-1}) \quad (4.7)$$

Darin sind x_u und x_o die Grenzen des Definitionsbereichs, L_x die Parameterlänge von x und $X_{G(y)}$ der Stellenwert der binären Kodierung von X an der Stelle y .

Eine Variante der Binärkodierung ist der Gray-Kode, der die in Kapitel 3.1.6 beschriebenen Vorteile bezüglich des Hamming-Abstands gegenüber einer standardbinären Kodierung besitzt. Die Berechnung der reellen Zahlen aus den Gray-kodierten erfolgt nach Gl. (4.8).

$$x_{dec} = x_u + \frac{x_o - x_u}{2^{L_x} - 1} \sum_{i=1}^{L_x} \left(X_{G(L_x-i)} (2^{L_x-i} - 1) (-1)^{\sum_{j=1}^i X_{G(L_x-j)}} \right) \quad (4.8)$$

Es sind x_u und x_o die Grenzen des Definitionsbereichs, L_x die Parameterlänge von x und $X_{G(y)}$ der Stellenwert der Gray-Kodierung von X an der Stelle y .

Bei der zufälligen Erzeugung der ersten Elterngeneration und auch bei den durchgeführten Rekombinationen und Mutationen können Parameterkombinationen erzeugt werden, die außerhalb der Definitionsbereiche liegen. Dies betrifft bei Mischverteilungen die Gewichtungparameter p_i und bei dreiparametrischen Weibullverteilungen in gemischten Verteilungen die ausfallfreien Zeiten t_{0i} .

Der Wertebereich für die Gewichtungparameter p_i ist das Intervall $[0, 1]$. Die Normierung der Summe der Gewichtungparameter p_i der Mischverteilungen auf den Wert eins erfolgt mit Gl. (4.9).

$$p_{i, Norm} = \frac{p_i}{\sum_{i=1}^N p_i} \quad (4.9)$$

Die ausfallfreien Zeiten t_{0i} sind nur im Intervall $(0, T_i]$ definiert, wodurch Parameterkombinationen mit $t_{0i} \geq T_i$ nicht zulässig sind. Die Wertebereiche der charakteristischen Lebensdauer (T_u, T_o) und der ausfallfreien Zeit (t_{0u}, t_{0o}) müssen zu Beginn jeder Parameterschätzung definiert werden. Diese Intervalle können sich überlappen, und dadurch können unzulässige Werte von $t_{0i} \geq T_i$ durch die genetischen Operatoren entstehen. Nach Gl. (4.10) berechnet sich der Wert $t_{0i, Norm}$ für den Wert $t_{0i} \geq T_i$.

$$t_{0i, Norm} = t_{0i} \frac{T_i - t_{0u}}{t_{0o} - t_{0u}} \quad (4.10)$$

Die so berechneten Werte werden als Binär- oder Gray-Kode in die Kodierung der Parameter geschrieben. Die Berechnung des Binärkodes aus reellen Zahlen erfolgt nach Gl. (4.11). Darin ist x_{bin} ein Vektor, der die Werte des Binärkodes der reellen Zahl x_{dec} enthält, L_x die Parameterlänge von x_{bin} , x_u und x_o die Grenzen des Definitionsbereichs und z ein Vektor, der in absteigender Reihenfolge die ganzzahligen Werte $L_x, \dots, 1$ enthält. Die Modulfunktion mod liefert den Rest der Division zweier Zahlen.

$$x_{bin} = \left(\frac{\begin{matrix} x_{dec} - x_u \\ x_o - x_u \end{matrix}}{2^z} 2^{L_x-1} \right) \bmod 2 \quad (4.11)$$

Für die Umwandlung von reellen Zahlen in eine Gray-Kode-Darstellung ist zuerst die Umwandlung in die standardbinäre Kodierung mit Gl. (4.11) erforderlich. Daraus berechnet sich mit Gl. (4.12) die Kodierung im Gray-Kode x_{Gray} .

$$x_{Gray} = |x_{bin} + x_{bin-1} - 2|(x_{bin} + x_{bin-1}) \quad (4.12)$$

Darin ist x_{bin} die standardbinäre Kodierung und x_{bin-1} eine um ein Bit nach rechts verschobene Variante von x_{bin} mit $x_{bin-1} = (0, x_{bin(L_x)}, x_{bin(L_x-1)}, \dots, x_{bin(2)}, x_{bin(1)})$.

Zur Evaluation des Einflusses der Kodierung auf die Güte der Parameterschätzung werden die Mischverteilungstypen aus Kapitel 4.1 verwendet. Es werden, um das statistische Rauschen durch die Monte-Carlo Methode auszuschließen, Mediandaten verwendet, die genau auf der Verteilungskurve liegen. Grundlage für den Vergleich der Güte der Schätzungen ist der erreichte minimale senkrechte Abstand der Datenpunkte zur Verteilungskurve, auf den im nachfolgenden Kapitel 4.3 detailliert eingegangen wird. Je geringer der Abstand, umso besser ist das gefundene Ergebnis.

Für die verschiedenen Parameter, die zur Initialisierung des Algorithmus und der Problemstellung erforderlich sind, wurden aus Voruntersuchungen gewonnene Parameter verwendet. Sofern nicht anders angegeben, werden die Initialisierungsdaten für die Untersuchungen in den nachfolgenden Kapiteln beibehalten.

In Tabelle 4.2 sind die für die Untersuchung des Einflusses der Kodierungsvariante genutzten Initialisierungsparameter des Algorithmus und der Problemstellung angegeben. Der Einfluss der einzelnen Algorithmenparameter auf die Schätzung der Verteilungsparameter wird ab Kapitel 4.4 beschrieben und diskutiert.

Tabelle 4.2: Initialisierungsparameter des Algorithmus und der Problemstellung

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Populationsgröße (E, N)	100,100	Eingangsdaten / Anzahl	Mediandaten / 100
Generationenanzahl	100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Fitnessfunktion	Vert. Abstand linear gew.		
Rekombination	1-Punkt-Crossover		
Mutationswahrscheinlichkeit	0,001		
Selektion	Stochastisch	Genauigkeit der Parameter	(3, 2, 0, 5)
	Eltern + Nachkommen	Anzahl Unterverteilungen	(2, 2)
	Fitnessselektion		

Es werden pro Variante und Verteilungstyp jeweils 20 Durchläufe des Algorithmus bzw. Replikationen durchgeführt. Da es sich bei Lösungen des genetischen Algorithmus meist um lokale Optima handelt, ist es erforderlich, mehrere Parameterschätzungen durchzuführen, um die Güte der Parameterkombination einschätzen zu können.

Zum Vergleich der Ergebnisse beider Kodierungsvarianten bei verschiedenen Mischverteilungstypen dienen Box-Plots, siehe Bild 4.6. Mittels Box-Plots ist es möglich, verschiedene statistische Kenngrößen übersichtlich darzustellen [118]. Die namensgebende Box umschließt das 50 %-Perzentil, im vorliegenden Fall zehn Werte. Die von der Box ausgehenden Linien, die sogenannten „Whisker“, zeigen den Bereich vom Minimal- bis zum Maximalwert. Zusätzlich sind der Median und der arithmetische Mittelwert eingetragen.

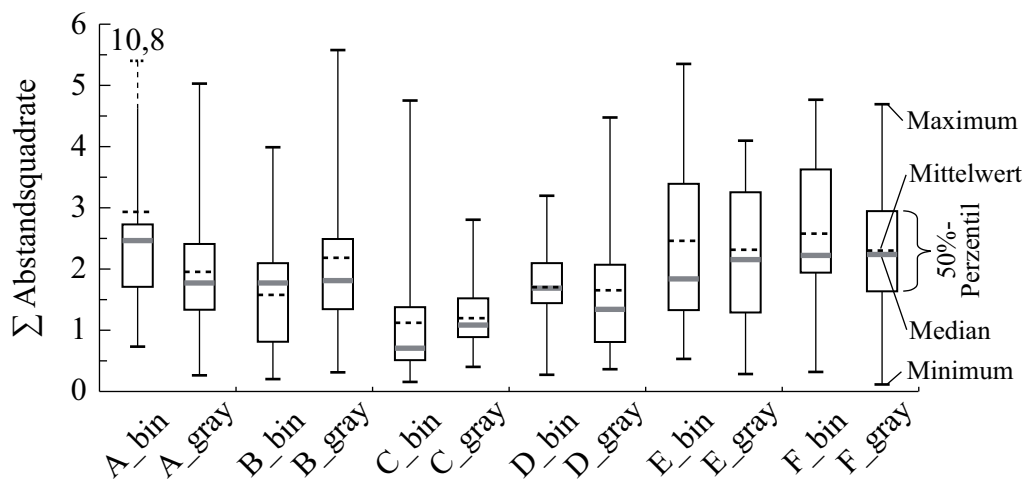


Bild 4.6: Vergleich der Fitnesswerte bei binärer und Gray-Kodierung für die Verteilungstypen A - F

Die Box-Plots in Bild 4.6 zeigen keine eindeutig erkennbare Tendenz. Die Streubereiche der 50 %-Perzentile sind beim Vergleich der Ergebnisse der standardbinären Kodierung und der Gray-Kodierung nahezu identisch. Die Summe der Abstandsquadrate streut im Bereich zwischen 0,5 und 3,5. Eine Verbesserung der Schätzungen durch die Gray-Kodierung, sowohl der Streubereiche als auch der Mittelwerte, ist bei den Typen A, E und F erkennbar. Die Werte der minimalen Abstandsquadrate liegen beim direkten Vergleich aller Typen auf einem ähnlichen Niveau.

Es scheint keine eindeutige Verschlechterung der Parameterschätzung durch die großen Hamming-Abstände bei Verwendung der standardbinären Kodierung zu geben. Die Rechenzeit bei Verwendung der Gray-Kodierung ist, durch die aufwändigere Berechnung und Rückrechnung, im Mittel um ca. 7,5 % höher. Aus diesem Grund wird im Folgenden nur die standardbinäre Kodierung weiterverwendet.

4.3 Definition der Fitnessfunktion

Die Definition der Fitness- oder Zielfunktion nimmt die zentrale Rolle bei der Verwendung von genetischen Algorithmen ein. Die Fitnessfunktion dient zur Bewertung der Güte der einzelnen Individuen und hat damit, abhängig von der Selektionsvariante, einen direkten Einfluss auf die Bildung der Nachkommengeneration.

Das Problem der Schätzung von Verteilungsparametern aus Ausfalldatenpunkten lässt sich auf eine Minimierung der Abstände aller Datenpunkte zur Verteilungskurve reduzieren. Die Verteilungskurve, die die minimalen Abstände zu den Datenpunkten aufweist, stellt die beste Variante dar. Es lassen sich bei der Minimierung der Abstände drei Möglichkeiten unterscheiden:

- Minimierung der Abstände in Ordinateurichtung
→ siehe Bild 4.7 links, Kapitel 4.3.1
- Minimierung der Abstände in Abszissenrichtung
→ siehe Bild 4.7 Mitte, Kapitel 4.3.2
- Minimierung der Abstandsquadrate
→ siehe Bild 4.7 rechts, Kapitel 4.3.3

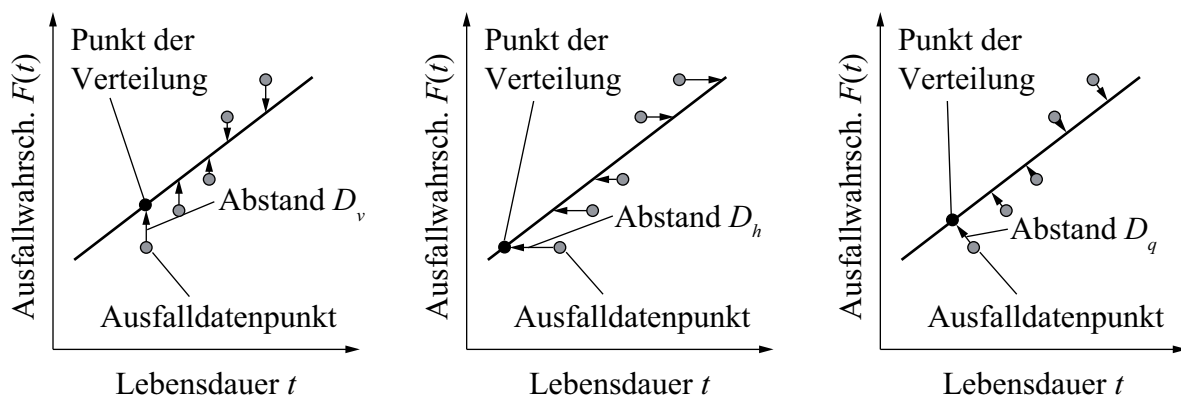


Bild 4.7: Möglichkeiten zur Abstandsminimierung

Eine weitere untersuchte Möglichkeit ist die Nutzung der Likelihoodfunktion Gl. (2.18) zum Vergleich der Individuen. Diese Möglichkeit wird in Kapitel 4.3.4 beschrieben.

4.3.1 Fitness in Ordinateurichtung

Die einfachste Variante der Fitnessberechnung ist die Bestimmung des Abstands der Datenpunkte zur gefundenen Verteilungskurve in Ordinateurichtung, siehe Bild 4.7 links. Der Fitnesswert für Mischverteilungen berechnet sich mit Gl. (4.13).

$$Z_{v,ges} = \sum_{i=1}^n |D_v| = \sum_{i=1}^n \left| F_{median}(t_i) - \sum_{j=1}^N (p_j F_j(t_i)) \right| \quad (4.13)$$

Für Serien- und Parallelsysteme ergeben sich die Ziel- oder Fitnessfunktionen Gl. (4.14) und Gl. (4.15).

$$Z_{v,ges} = \sum_{i=1}^n |D_v| = \sum_{i=1}^n \left| F_{median}(t_i) - \left(1 - \prod_{j=1}^N (1 - F_j(t))\right) \right| \quad (4.14)$$

$$Z_{v,ges} = \sum_{i=1}^n |D_v| = \sum_{i=1}^n \left| F_{median}(t_i) - \prod_{j=1}^N F_j(t) \right| \quad (4.15)$$

Der Vorteil dieser Variante besteht in der einfachen Berechnung der Ausfallwahrscheinlichkeiten $F(t)$. Zudem können die Werte der Ausfallwahrscheinlichkeiten zur graphischen Darstellung der Datenpunkte ebenfalls genutzt werden.

Beispielhaft ist in Bild 4.8 die Fitnessfunktion für die Berechnung der Fitness in Ordinateurichtung dargestellt.

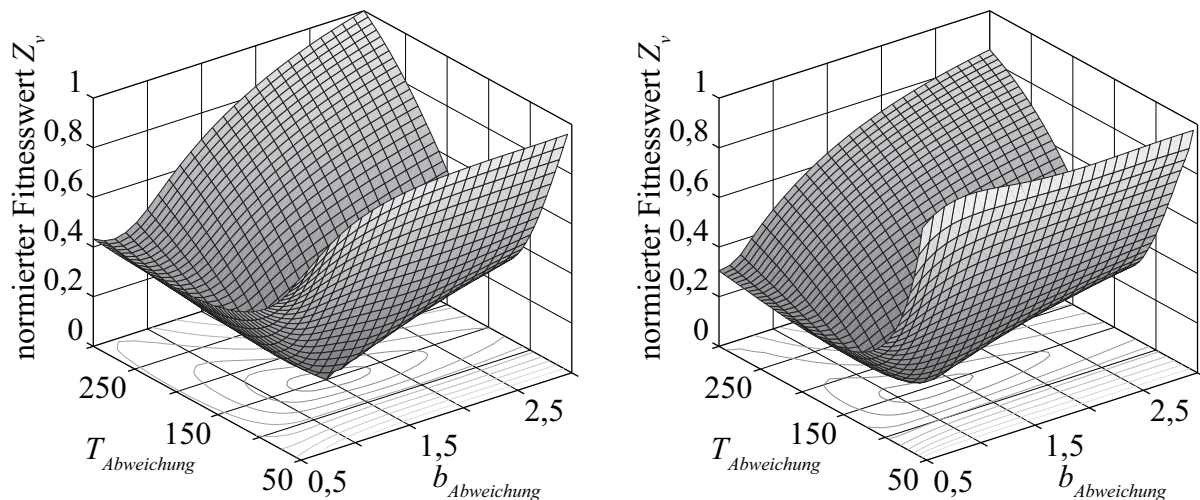


Bild 4.8: Darstellung der Fitnessfunktion für eine Weibullverteilung bei Bestimmung der Fitness in Ordinateurichtung ohne t_0 (links), mit t_0 (rechts)

Aufgetragen ist der normierte Fitnesswert über den Parametern der geschätzten zweiparametrischen Weibullverteilung mit den Parametern $b_{Abweichung}$ und $T_{Abweichung}$. Grundlage ist eine Weibullverteilung mit $b = 1,5$ und $T = 150$, Bild 4.8 links. Erkennbar ist in Bild 4.8 links das klar definierte Optimum des Fitnesswertes bei $b_{Abweichung} = 1,5$ und $T_{Abweichung} = 150$ mit dem Fitnesswert Null. In alle Richtungen steigt die Fitness stark an. Die höchsten Gradienten finden sich bei der Variation von $T_{Abweichung}$, wohingegen eine Abweichung des Formparameters $b_{Abweichung}$ für $b_{Abweichung} > b$ einen nur schwachen Gradienten bewirkt. Hohe Gradienten begünstigen eine schnelle Konvergenz des genetischen Algorithmus, da sich benachbarte Lösungsvarianten eindeutig durch ihre Fitnesswerte unterscheiden lassen.

Die Grundlage für Bild 4.8 rechts bildet eine Weibullverteilung mit den Parametern $b = 1,5$, $T = 150$ und $t_0 = 50$. Aufgetragen ist der normierte Fitnesswert über den

Parametern $b_{\text{Abweichung}}$ und $T_{\text{Abweichung}}$. Der Parameter der ausfallfreien Zeit ist konstant mit $t_{0\text{Abweichung}} = 35$. Der Verlauf der Fitnessfunktion mit t_0 ist im Vergleich mit jener ohne t_0 annähernd identisch. Das Optimum des Fitnesswertes liegt auf Grund von $t_{0\text{Abweichung}} \neq t_0$ mit 0,029 über Null und ist leicht zu einem kleineren Formparameter b verschoben. Der Wert der charakteristischen Lebensdauer für das Optimum liegt bei $T_{\text{Abweichung}} = 150$. Eine leichte Abweichung von der gesuchten charakteristischen Lebensdauer ergibt eine deutliche Erhöhung des Fitnesswertes. Dies führt dazu, dass diese Art der Fitnessbestimmung gut geeignet ist die charakteristische Lebensdauer zu schätzen.

Bei der Parameterschätzung von dreiparametrischen Weibullverteilungen mit der obigen Fitnessfunktion und 50 Datenpunkten ergibt sich ein großer Streubereich der geschätzten Verteilungskurven, siehe Bild 4.9 links. Die Initialisierungsparameter des Algorithmus und der Problemstellung sind im Anhang in Tabelle A1.1 aufgeführt. Die Box-Plots in Bild 4.9 zeigen die Abweichungen der geschätzten Verteilungskurven sowie die Abweichungen der Parameter b , T und t_0 . Die Darstellung der Box-Plots entspricht der Darstellung in Bild 4.6.

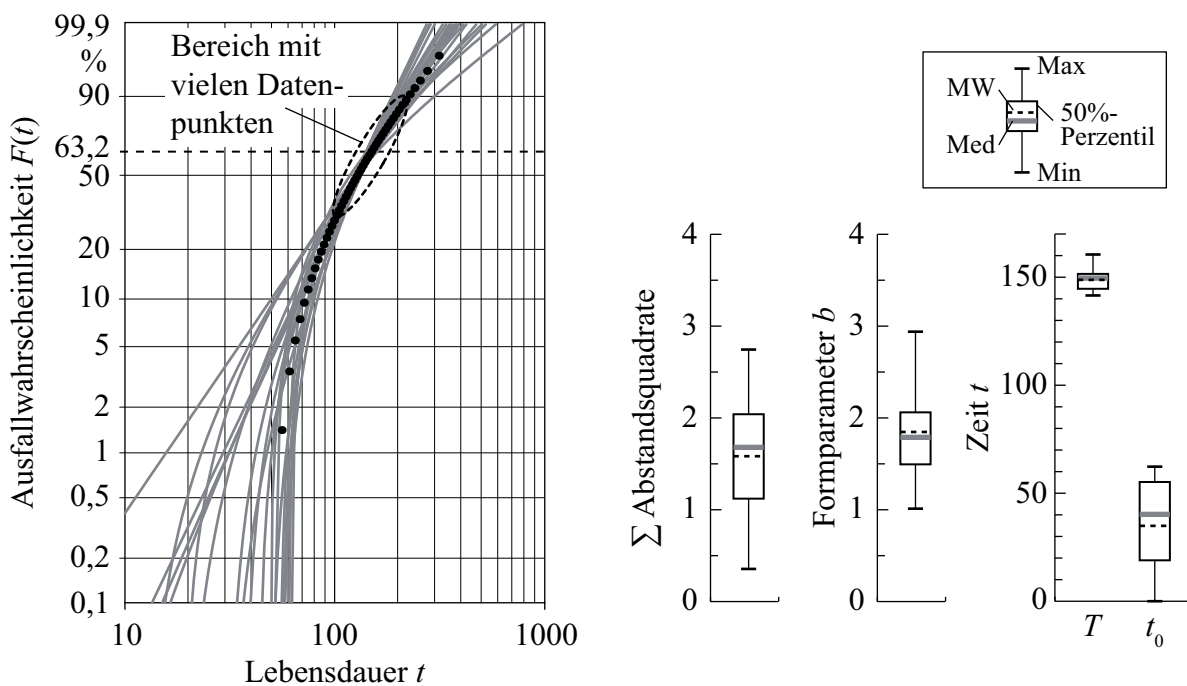


Bild 4.9: Parameterschätzungen für dreiparametrische Weibullverteilung mit $b = 1,5$, $T = 150$ und $t_0 = 50$ mit Fitness in Ordinateurichtung

Auffällig ist die geringe Streuung der charakteristischen Lebensdauer T und die im Vergleich dazu große Streuung der ausfallfreien Zeit t_0 . Das 50 %-Perzentil von t_0 ist 5,2-mal größer als das von T . Durch die große Dichte an Datenpunkten im Bereich von 63,2 %, siehe Bild 4.9, wirkt sich eine Verbesserung der Schätzung von T sehr viel stärker auf die Fitness aus als die Schätzung von t_0 . Dieser Bereich beinhaltet weit

weniger Datenpunkte, wodurch bei der Selektion Individuen mit guter Schätzung von T bevorzugt werden.

Die durchschnittliche Rechenzeit für eine Replikation mit 100 Generationen bei einer Populationsgröße von 100 Eltern und 100 Nachkommen beträgt etwa 4 s. Zum Einsatz kam dazu ein Rechner mit einem Intel[®] Xeon[®] W3680 mit 3,33 GHz und 6 GB Hauptspeicher mit einem Windows 7[®] Betriebssystem von Microsoft[®].

4.3.2 Fitness in Abszissenrichtung

Die zweite Variante der Fitnessberechnung über die Abstände der Datenpunkte zur Verteilungskurve ist die Berechnung der Fitness in Abszissenrichtung, siehe Bild 4.7 Mitte. Die Berechnung der Fitness für Mischverteilungen erfolgt nach Gl. (4.16).

$$Z_{h,ges} = \sum_{i=1}^n |D_h| = \sum_{i=1}^n |t_i - F_M^{-1}(F_{median}(t_i))| \quad (4.16)$$

Für Serien- und Parallelsysteme muss anstatt der Inversen der Mischverteilung F_M^{-1} die Inverse der Systemgleichung F_S^{-1} verwendet werden.

$$Z_{h,ges} = \sum_{i=1}^n |D_h| = \sum_{i=1}^n |t_i - F_S^{-1}(F_{median}(t_i))| \quad (4.17)$$

Da die Verteilungskurven bei gemischten Weibullverteilungen mit ausfallfreier Zeit nur in Intervallen definiert sind, kann die Inverse nicht direkt berechnet werden. Die Berechnung der Zeiten zur Fitnessbestimmung erfolgt über die Bisektionsmethode.

In Bild 4.10 ist der Verlauf der Fitnessfunktion für die Fitness in Abszissenrichtung dargestellt.

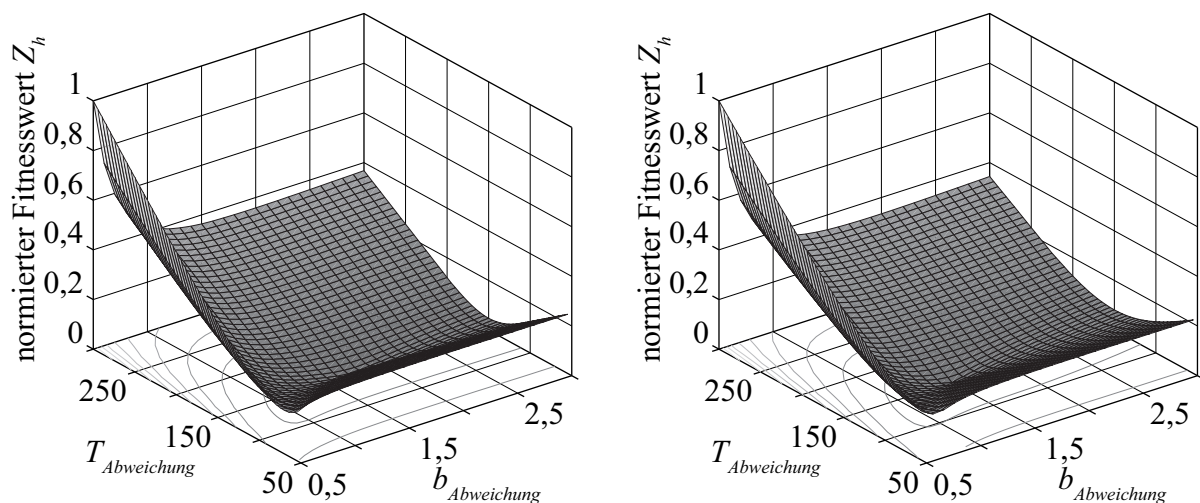


Bild 4.10: Darstellung der Fitnessfunktion für eine Weibullverteilung bei Bestimmung der Fitness in Abszissenrichtung ohne t_0 (links), mit t_0 (rechts)

Aufgetragen ist der normierte Fitnesswert Z_h über den Parametern der geschätzten zweiparametrischen Weibullverteilung mit den Parametern $b_{Abweichung}$ und $T_{Abweichung}$. Die Parameter der zu Grunde liegenden Weibullverteilung sind mit den in Kapitel 4.3.1 verwendeten identisch.

Im Gegensatz zur Fitnessbestimmung in Ordinatenrichtung ist das Optimum der dargestellten Fitnessfunktion nur schwach ausgeprägt. Der größte Gradient tritt bei der Verringerung von $b_{Abweichung}$ auf. Die Variation von $T_{Abweichung}$, bei gleichbleibendem Formparameter, hat nur eine geringe Änderung der Fitness zur Folge. Durch das schwach ausgeprägte Optimum ist es schwieriger als bei der Variante in Ordinatenrichtung die Parameter genau zu schätzen. In Bild 4.11 sind die Ergebnisse der Parameterschätzung mit der Fitnessberechnung in Abszissenrichtung dargestellt.

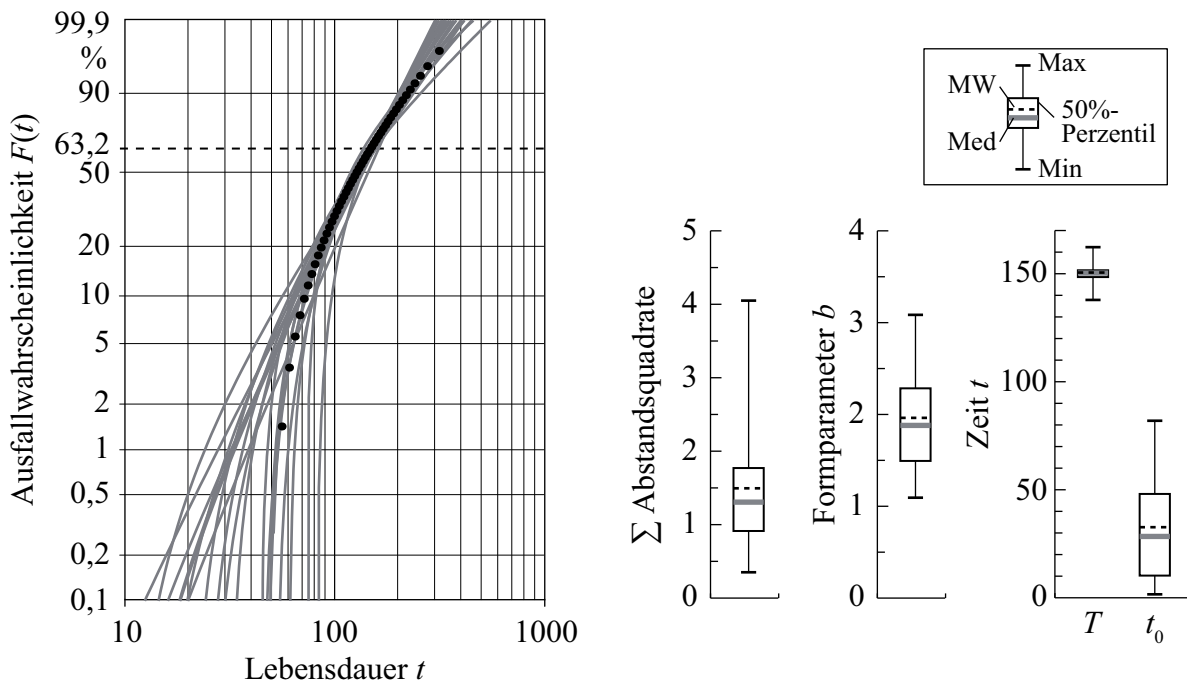


Bild 4.11: Parameterschätzungen für dreiparametrische Weibullverteilung mit $b = 1,5$, $T = 150$ und $t_0 = 50$ mit Fitness in Abszissenrichtung

Das Ergebnis der Summe der Abstandsquadrate ist vergleichbar mit den Werten, die mit der Fitnessbestimmung in Ordinatenrichtung erreicht werden. Das 50 %-Perzentil der Streuung des Formparameters ist etwa 1,5-mal größer als die mit der Fitnessbestimmung in Ordinatenrichtung erreichte. Auffällig ist die sehr gute Schätzung der charakteristischen Lebensdauer mit einer sehr geringen Streuung. Dagegen ist die Streuung der ausfallfreien Zeit ähnlich groß wie bei der Variante aus Kapitel 4.3.1.

Der Nachteil der Methode ist die aufwändige iterative Berechnung der Fitnesswerte über die Bisektionsmethode. Dadurch steigt die mittlere Berechnungszeit pro Replikation von 4 s für die ordinatenbasierte Berechnung auf 153 s für die abszissenbasierte Berechnung.

4.3.3 Fitness mit der Methode der minimalen Abstandsquadrate

Die dritte Variante zur Berechnung der Fitness mit Hilfe der Abstände der Datenpunkte von der Verteilungskurve ist in Bild 4.7 rechts dargestellt. Die Anwendung dieser Methode zur Schätzung der Gewichtungsfaktoren wurde von Woodward et al in [119] diskutiert. Hierbei wird der Abstand nicht nur in Ordinaten- oder Abszissenrichtung betrachtet, sondern der minimale Abstand der Datenpunkte zur Verteilungskurve. Der minimale Abstand berechnet sich für Mischverteilungen nach Gl. (4.18) und für Serien- und Parallelsysteme nach Gl. (4.19).

$$Z_{q,ges} = \sum_{i=1}^n \sqrt{D_q} = \sum_{i=1}^n \sqrt{D_t^2 + D_F^2} = \sum_{i=1}^n \sqrt{(t_v - t_i)^2 + (F_M(t_v) - F_{median}(t_i))^2} \quad (4.18)$$

$$Z_{q,ges} = \sum_{i=1}^n \sqrt{D_q} = \sum_{i=1}^n \sqrt{D_t^2 + D_F^2} = \sum_{i=1}^n \sqrt{(t_v - t_i)^2 + (F_S(t_v) - F_{median}(t_i))^2} \quad (4.19)$$

Darin sind t_v und $F(t_v)$ die Abszissen- und Ordinatenwerte des am nächsten zu $(t_i, F(t_i))$ liegenden Punktes auf der Verteilungskurve. Aufgrund der Monotonie kann der Ausdruck unter der Wurzel minimiert werden [120], was zu Punkt $(t_v, F(t_v))$ führt. Damit ergibt sich für Mischverteilungen die Lösung aus der Nullstelle aus Gl. (4.20).

$$\frac{dD_q}{dt_v} = 2(t_v - t_i) + 2f_M(t_v)(F_M(t_v) - F_{median}(t_i)) = 0 \quad (4.20)$$

Die Lösung von Gl. (4.20) wird iterativ über die Bisektionsmethode bestimmt. Das Vorgehen für boolesche Systeme ist identisch, wegen der Produkte aber aufwändiger.

Bild 4.12 zeigt den Verlauf der Fitnessfunktion für die mittels minimaler Abstandsquadrate bestimmte Fitness.

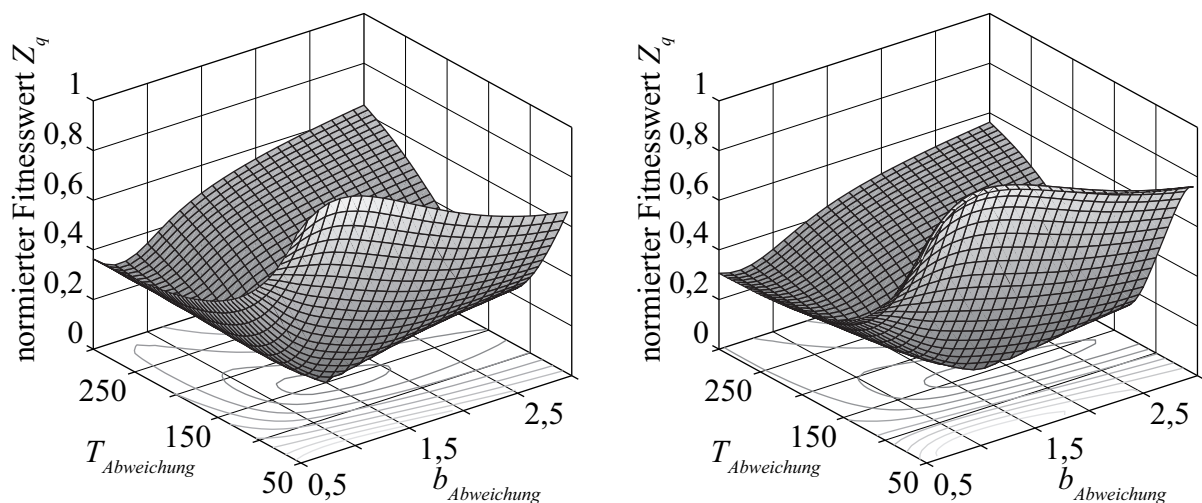


Bild 4.12: Darstellung der Fitnessfunktion für eine Weibullverteilung bei Bestimmung der Fitness mittels minimaler Abstandsquadrate ohne t_0 (links), mit t_0 (rechts)

Aufgetragen ist der normierte Fitnesswert Z_q über den Parametern der geschätzten zweiparametrischen Weibullverteilung mit den Parametern $b_{\text{Abweichung}}$ und $T_{\text{Abweichung}}$. Die verwendeten Parameter der zu Grunde liegenden Weibullverteilung sind mit den in Kapitel 4.3.1 verwendeten identisch.

Die Fitnessfunktionen ähneln sehr stark den Fitnessfunktionen, die sich aus der Bestimmung der Fitness in Ordinateurichtung ergeben. Dies ist in den sehr unterschiedlichen Skalierungen der Ordinateur- und Abszissenachsen begründet. Eine Abweichung auf der Abszissenachse bewirkt eine größere Abstandszunahme als die Abweichung auf der Ordinateurachse. Dadurch berechnet sich der minimale Abstand einer Verteilungskurve zu einem gegebenen Punkt meist direkt aus der Differenz der Ordinateurwerte. Die Unterschiede der Fitnesswerte ergeben sich damit aus den Fitnesswerten bei kleinen Ausfallzeiten. Eine Möglichkeit dies zu verbessern besteht in der Normierung der Zeitachse, wodurch die Unterschiede der Werte der Ordinateur- und Abszissenachse egalisiert würden. Die Unterschiede der Fitnessfunktion bei einer vorhandenen ausfallfreien Zeit sind deutlicher. Das Minimum der Fitness liegt bei einem Wert von $b > 1,5$ und ist nicht so deutlich ausgeprägt wie bei der Fitnessbestimmung in Ordinateurichtung.

Die Ergebnisse der Parameterschätzung mit minimalen Abstandsquadraten sind in Bild 4.13 dargestellt.

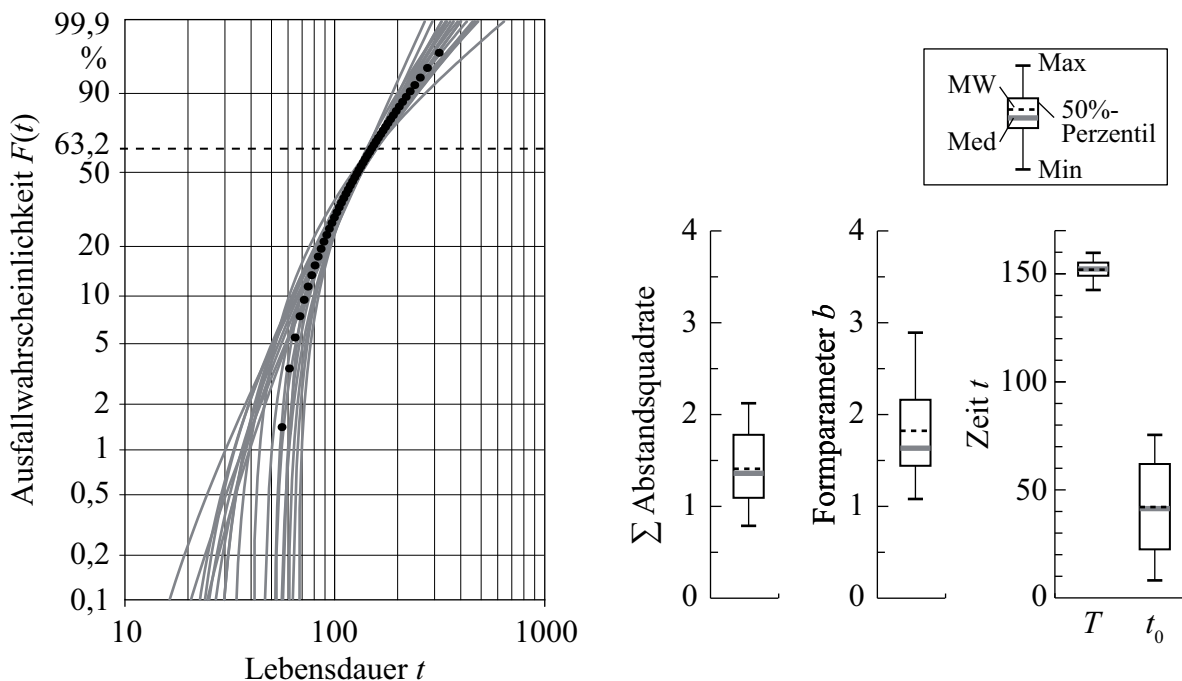


Bild 4.13: Parameterschätzungen für dreiparametrische Weibullverteilung mit $b = 1,5$, $T = 150$ und $t_0 = 50$ mit Fitnessberechnung mittels minimalen Abstandsquadraten

Die Schätzungen sind vergleichbar mit denen der Fitnessbestimmung in Ordinateurichtung, siehe Bild 4.9. Die Ähnlichkeit der Ergebnisse ist auf die sehr ähnlichen

Verläufe der Fitnessfunktionen zurückzuführen. Auch bei dieser Variante ist die Schätzung der charakteristischen Lebensdauer sehr gut. Der Streubereich des 50 %-Perzentsils ist mit denen der vorherigen Methoden vergleichbar. Die Schätzung der ausfallfreien Zeit zeigt bei dieser Methode die besten Ergebnisse. Das Vergleichskriterium Summe der Abstandsquadrate zeigt nur eine geringe Verbesserung gegenüber den vorigen Methoden, obwohl die betrachtete Methode auf diesem Kriterium basiert.

Wie bei der Variante der Fitnessberechnung in Abszissenrichtung ist durch die Bisektionsmethode zur Bestimmung des minimalen Abstands die Rechenzeit mit 249 s pro Replikation sehr lang. Da für die Berechnung sowohl die Funktionswerte selbst als auch deren Ableitungen berechnet werden müssen, steigt der Rechenzeitbedarf gegenüber der abszissenbasierten Variante zusätzlich an.

4.3.4 Likelihoodfunktion als Fitnessfunktion

Eine Variante zur Berechnung der Fitness unabhängig vom Abstand der Verteilungskurve zu den Datenpunkten ist die Likelihoodfunktion Gl. (2.18). Durch die Verwendung der Likelihoodfunktion wird das bisherige Minimierungsproblem der Abstände ein Maximierungsproblem des Likelihoodfunktionwertes.

Der Verlauf der Likelihoodfunktion ist in Bild 4.14 dargestellt. Aufgetragen ist der normierte Fitnesswert Z_L über den Parametern der geschätzten zweiparametrischen Weibullverteilung mit den Parametern $b_{\text{Abweichung}}$ und $T_{\text{Abweichung}}$. Die Parameter der zu Grunde liegenden Weibullverteilung sind mit den in Kapitel 4.3.1 verwendeten identisch.

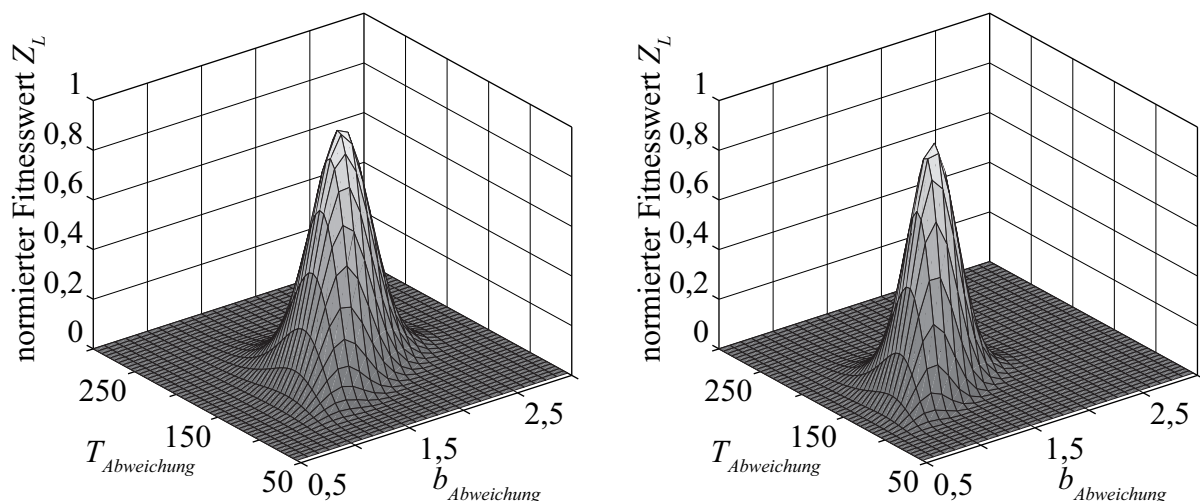


Bild 4.14: Darstellung der Fitnessfunktion für eine Weibullverteilung bei Bestimmung der Fitness über die Likelihoodfunktion ohne t_0 (links), mit t_0 (rechts)

Der Verlauf der Likelihoodfunktion ohne ausfallfreie Zeit, Bild 4.14 links, zeigt ein Maximum bei den Parametern $b_{\text{Abweichung}} = 1,5$ und $T_{\text{Abweichung}} = 150$. Mit ausfallfreier

Zeit und einem geschätzten $t_0 = 35$ ergibt sich ein leicht zu kleineren Werten von $b_{\text{Abweichung}}$ verschobenes Maximum. Für die Optimierung ergibt sich mit dem scharf abgegrenzten Optimum eine gute Möglichkeit der Verwendung als Fitnessfunktion.

Für das Testproblem mit $b = 1,5$, $T = 150$ und $t_0 = 50$ wurden die Initialisierungsparameter des Algorithmus und der Problemstellung aus Tabelle A1.1 im Anhang verwendet. Durch die Festlegung des Intervalls $[0, 100]$ für t_0 konnte kein stabiler Lauf des genetischen Algorithmus erreicht werden. Die Berechnung der Likelihoodfunktion für $t_{i,\min} < t_0$ führt zu nicht definierten Lösungen. Die Angabe eines so großen Intervalls ist jedoch für die Parameterschätzungen von Mischverteilungen mit ausfallfreier Zeit erforderlich, da für alle Verteilungen ein gemeinsamer Lösungsraum definiert wird. Die Likelihoodfunktion als Fitnessfunktion ist somit in der untersuchten Variante nicht geeignet, weshalb im Folgenden nur noch die Varianten aus Kapitel 4.3.1 bis Kapitel 4.3.3 weiterverfolgt werden.

4.3.5 Vergleich der Varianten zur Fitnessbestimmung

In den vorhergehenden Kapiteln wurden die verschiedenen Möglichkeiten zur Bestimmung der Fitness der Individuen vorgestellt. Bild 4.15 zeigt den direkten Vergleich der erzielten Ergebnisse.

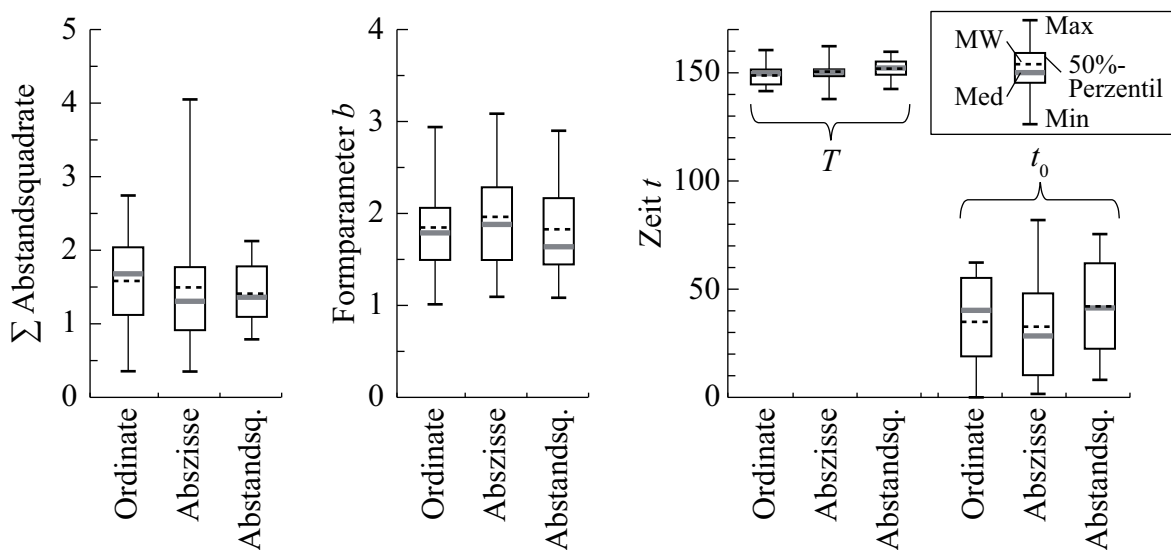


Bild 4.15: Vergleich der Varianten zur Fitnessberechnung

Im Allgemeinen zeigen die verschiedenen Varianten nur geringe Unterschiede. Alle Varianten zeigen bei der Schätzung der charakteristischen Lebensdauer sehr gute Ergebnisse, was die Bedeutung dieses Wertes für die Güte der Schätzung unterstreicht. Dagegen streut die Bestimmung der ausfallfreien Zeit bei allen Varianten stark. Im direkten Vergleich schneidet die Berechnung der Fitness über die Ordinatenwerte etwas schlechter ab als die beiden anderen Varianten. Die Gesamtstreuung liegt bei der

Berechnung über die Abszissenwerte höher als bei den beiden anderen Verfahren. Die Streuung des 50 %-Perzents zeigt diese Eigenschaft nicht. Leichte Vorteile bezüglich der Summe der Abstandsquadrate und der Schätzung der ausfallfreien Zeit t_0 zeigt die Berechnung der Fitness mittels der minimalen Abstandsquadrate.

Die Rechenzeit bei der Bestimmung der Fitness über die Ordinatenwerte liegt bei etwa 4 s pro Replikation und 100 Generationen mit jeweils 100 Eltern und Nachkommen. Die Varianten mit der Fitnessberechnung in Abszissenrichtung und der Fitnessberechnung mittels minimalen Abstandsquadraten benötigen mit 153 s und 249 s deutlich mehr Rechenzeit. Mit der Zunahme an Rechenzeit geht gleichzeitig keine signifikante Verbesserung der Genauigkeit der geschätzten Parameter einher. Die geringe Verbesserung rechtfertigt den vielfachen Zeitaufwand nicht. Deshalb wird nachfolgend die Berechnung der Fitness in Ordinatenrichtung durchgeführt.

4.3.6 Gewichtung der Datenpunkte

Die Schätzung der ausfallfreien Zeit ist für die Zuverlässigkeitstechnik von großer Bedeutung, da erst ab diesem Zeitpunkt mit Ausfällen gerechnet werden muss. Zusätzlich bedeutet eine schlechte Approximation der Werte in Bereichen geringer Ausfallwahrscheinlichkeit eine schlechte Schätzung der Parameter einer Unterverteilung, die nur in diesen Bereichen dominiert. Bild 4.16 zeigt die Schätzungen der Mischverteilung Typ E bei allen 20 Replikationen.

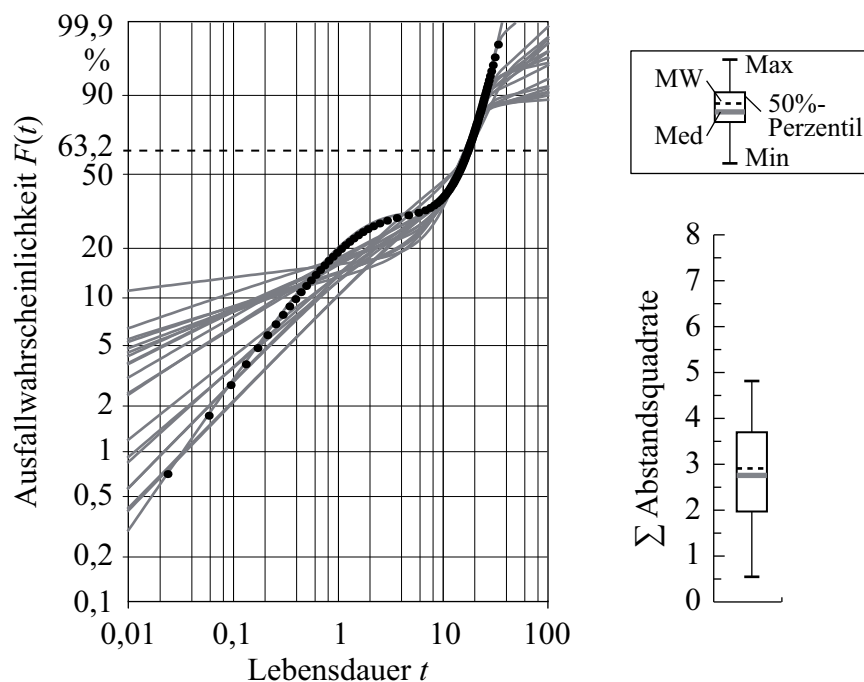


Bild 4.16: Ergebnisse der Parameterschätzungen der Mischverteilung Typ E

Die Mischverteilung Typ E zeigt in Bild 4.6 den größten Streubereich und das größte 50 %-Perzentil, was darauf hindeutet, dass die Parameter dieses Typs schwer zu

schätzen sind. Deshalb wird für die folgenden Untersuchungen Typ E verwendet. Die Parameter des genetischen Algorithmus finden sich im Anhang in Tabelle A1.2.

In Bild 4.16 ist zu erkennen, dass eine Annäherung an die Datenpunkte in Bereichen mit geringer und hoher Ausfallwahrscheinlichkeit ungenügend ist. Im Bereich zwischen 50 % und 80 % Ausfallwahrscheinlichkeit ist die Schätzung bei nahezu allen Replikationen sehr gut. Insgesamt bedeutet dies eine schlechte Abbildung der Datenpunkte durch die geschätzten Verteilungen.

Zur Verbesserung der Approximation der Datenpunkte im Bereich geringer und hoher Ausfallwahrscheinlichkeiten werden im Folgenden zwei Möglichkeiten vorgestellt und diskutiert.

Lineare Gewichtung

Durch die Einführung einer Gewichtung der Datenpunkte ist es möglich die Güte der Schätzung in Bereichen mit geringer und hoher Ausfallwahrscheinlichkeit zu verbessern. Eine Möglichkeit ist eine lineare Gewichtung der vorhandenen Datenpunkte. Grundlage dafür ist der Abstand a_i zwischen zwei benachbarten Datenpunkten. Der Gewichtungsvektor \underline{w} ist durch Gl. (4.21) definiert.

$$\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \\ w_n \end{bmatrix} = \begin{bmatrix} a_1 \\ 0,5(a_1 + a_2) \\ \vdots \\ 0,5(a_{n-1} + a_n) \\ a_{n-1} \end{bmatrix} \quad (4.21)$$

Bei der Berechnung des Fitnesswertes wird der Gewichtungsvektor mit den jeweiligen Ordinatenabständen D_v multipliziert. Daraus ergibt sich mit der um den Gewichtungsvektor \underline{w} erweiterten Gl. (4.13) für Mischverteilungen Gl. (4.22). Die Erweiterung der Gleichungen für die Serien- und Parallelsysteme erfolgt analog.

$$Z_{v,ges,lin} = \sum_{i=1}^n w_i |D_v| = \sum_{i=1}^n w_i \left| F_{median}(t_i) - \sum_{j=1}^N (p_j F_j(t_i)) \right| \quad (4.22)$$

Die Gewichtung bewirkt eine Anhebung der Abweichung in Bereichen mit wenigen Datenpunkten, was große Abstände zwischen den Punkten bedeutet. Dadurch erfahren Individuen mit einer guten Approximation dieser Bereiche eine Erhöhung der Selektionswahrscheinlichkeit. Die Auswirkung auf die Bereiche mit vielen, eng beieinander liegenden Datenpunkten ist mit einer Straffunktion oder Penaltyfunktion vergleichbar, da in diesen Bereichen die Gewichtung einen Malus für die Fitnessberechnung bewirkt.

In Bild 4.17 ist die Schätzung der Parameter derselben Verteilung unter Einbeziehung des Gewichtungsvektors \underline{w} dargestellt. Gegenüber den geschätzten Verteilungen ohne Gewichtung in Bild 4.16 ist keine Verringerung der Streuung der Verteilungen im Bereich geringer Ausfallwahrscheinlichkeiten erkennbar.

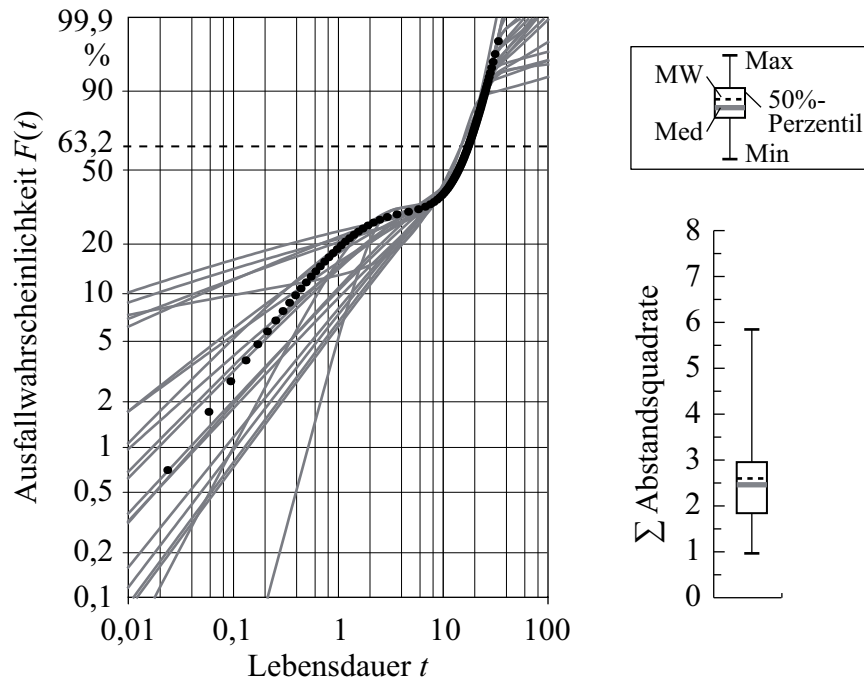


Bild 4.17: Ergebnisse der Parameterschätzungen der Mischverteilung Typ E mit Gewichtungsvektor \underline{w} linear gewichtet

Dagegen vergrößert sich der Bereich, in dem die Datenpunkte gut angenähert werden, auf einen Bereich von 35 % bis 90 % Ausfallwahrscheinlichkeit. Die Streuung im Bereich hoher Ausfallwahrscheinlichkeiten ist etwas geringer. Der Vergleich der Box-Plots der Schätzung mit Gewichtungsvektor \underline{w} zeigt eine Verkleinerung des 50 %-Perzentils und gleichzeitig eine Vergrößerung der Gesamtstreuung gegenüber der Schätzung ohne \underline{w} . Der Median- und Mittelwert sind etwas geringer.

Die Einführung der Gewichtung bewirkt somit eine leichte Verbesserung der Schätzung im Bereich mittlerer Ausfallwahrscheinlichkeiten. Die Streuung der geschätzten Verteilungen ist jedoch nicht geringer. Der linear gewichtete Fitnesswert als Messgröße für die Abweichung kann nur noch zum Vergleich der Individuen verwendet werden. Der Wert an sich lässt durch die Gewichtung keine direkten Rückschlüsse auf die Abweichung von den Datenpunkten zu.

Gewichtung in Weibullkoordinaten

Eine weitere mögliche Variante der Gewichtung ist eine Gewichtung bezogen auf die Skalierung des Weibullwahrscheinlichkeitspapiers. Die Skalierung des Weibullwahrscheinlichkeitsnetzes bewirkt, dass Bereiche mit geringer und hoher

Ausfallwahrscheinlichkeit gestreckt und die dazwischenliegenden Bereiche gestaucht werden. Die Transformationen der Ordinate $F'(t)$ und der Abszisse t' sind in Gl. (4.23) und Gl. (4.24) gezeigt.

$$F'(t) = \ln(-\ln(1 - F(t))) \quad (4.23)$$

$$t' = \ln t \quad (4.24)$$

Übertragen auf die Abstände zwischen den einzelnen Datenpunkten bedeutet das

$$a_i = \sqrt{(\ln t_{i+1} - \ln t_i)^2 + (\ln(-\ln(1 - F(t_{i+1}))) - \ln(-\ln(1 - F(t_i))))^2}. \quad (4.25)$$

Der Gewichtungsvektor \underline{w} berechnet sich, analog zur linearen Gewichtung, aus den Abständen der Datenpunkte a_i mit Gl. (4.21). Die Gewichtungen der Datenpunkte der zur Evaluation verwendeten Weibullverteilung sind in Bild 4.18 für die lineare Gewichtung und die Gewichtung in Weibullkoordinaten dargestellt.

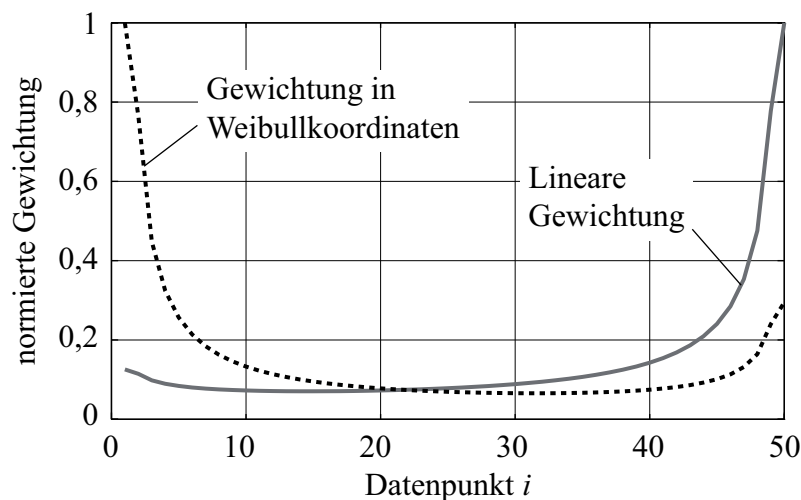


Bild 4.18: Verläufe der linearen Gewichtung und der Gewichtung in Weibullkoordinaten

Die Gewichtungen zeigen einen gegensätzlichen Verlauf. Die lineare Gewichtung erhöht den Fitnesswert für Datenpunkte mit niedrigem Rang, also im Bereich geringer Ausfallwahrscheinlichkeiten, nur leicht. Dagegen wird der Bereich hoher Ausfallwahrscheinlichkeiten sehr stark gewichtet, womit sich eine Verbesserung der Schätzung der Verteilung im Bereich hoher Ausfallwahrscheinlichkeiten erklären lässt.

Die Gewichtung in Weibullkoordinaten zeigt einen entgegengesetzten Verlauf. Abweichungen im Bereich niedriger Ausfallwahrscheinlichkeiten werden sehr stark gewichtet, wodurch eine Abweichung in diesem Bereich eine starke Verschlechterung der Fitness und damit der Selektionswahrscheinlichkeit bewirkt. Dadurch werden beispielsweise Individuen mit einer guten Schätzung der ausfallfreien Zeit bevorzugt selektiert. Im Bereich hoher Ausfallwahrscheinlichkeiten ist ebenfalls eine leichte

Erhöhung der Gewichtung erkennbar, der mittlere Bereich ist bei beiden Varianten nahezu gleich gewichtet.

In Bild 4.19 sind die Ergebnisse der Schätzung der Verteilung mit einer Gewichtung in Weibullkoordinaten dargestellt.

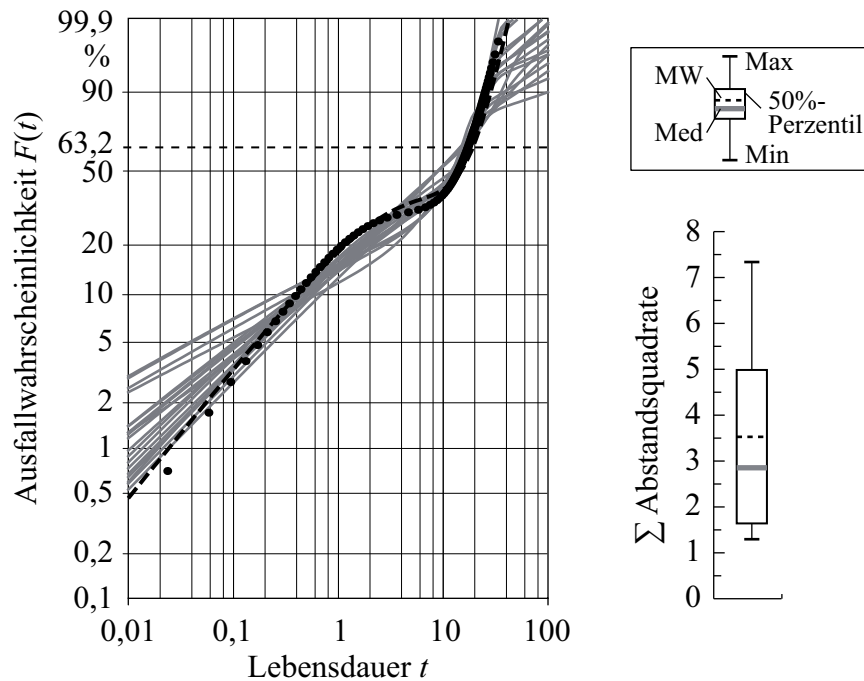


Bild 4.19: Ergebnisse der Parameterschätzungen der Mischverteilung Typ E mit Gewichtungsvektor \underline{w} in Weibullkoordinaten gewichtet

Im Weibulldiagramm ist eine Verringerung der Streuung bei geringen Ausfallwahrscheinlichkeiten zu erkennen. Dies deutet auf die Wirksamkeit der Gewichtung der Datenpunkte im Bereich geringer Ausfallwahrscheinlichkeiten hin. Insgesamt ist der Streubereich der Lösungsvarianten schmäler als bei der ungewichteten oder linear gewichteten Variante. Eine Einschnürung der geschätzten Verteilungen im mittleren Ausfallwahrscheinlichkeitsbereich wie in Bild 4.16 und Bild 4.17 ist nicht mehr erkennbar. Zusätzlich ist in Bild 4.19 die Verteilungskurve der besten Parameterschätzung als gestrichelte Linie dargestellt. Die Approximation der Daten im unteren Ausfallwahrscheinlichkeitsbereich ist gut. Im Bereich von $t = 2,5$ bis 10 liegt die Verteilungskurve über den zu approximierenden Datenpunkten. Im Bereich oberhalb von $t = 10$ liegt wiederum eine gute Approximation der Daten vor.

Die Streuung der Summe der Abstandsquadrate in Bild 4.19 zeigt eine Vergrößerung des gesamten Streubereichs und auch des 50 %-Perzentils. Die Erhöhung der Streuung rührt aus der absichtlichen Verschlechterung der Anpassung an die Datenpunkte im mittleren Ausfallwahrscheinlichkeitsbereich durch die Gewichtung her. Der Median der Schätzungen liegt ungefähr auf dem Niveau der ungewichteten Variante.

Wie bei der linearen Gewichtung verliert der Fitnesswert durch die Gewichtung in Weibullkoordinaten seine Aussagefähigkeit bezüglich der Abweichung der Lösungsvariante von den Datenpunkten. Es sind nur noch Vergleiche der Individuen untereinander möglich.

Zum Vergleich werden die Gewichtungsvarianten auf die Parameterschätzung der in Kapitel 4.3.1 verwendeten dreiparametrischen Weibullverteilung mit $b = 1,5$, $T = 150$ und $t_0 = 50$ angewandt. Bild 4.20 zeigt den Vergleich der Auswirkungen der verschiedenen Gewichtungen auf die Parameterschätzung der einzelnen Weibullverteilung.

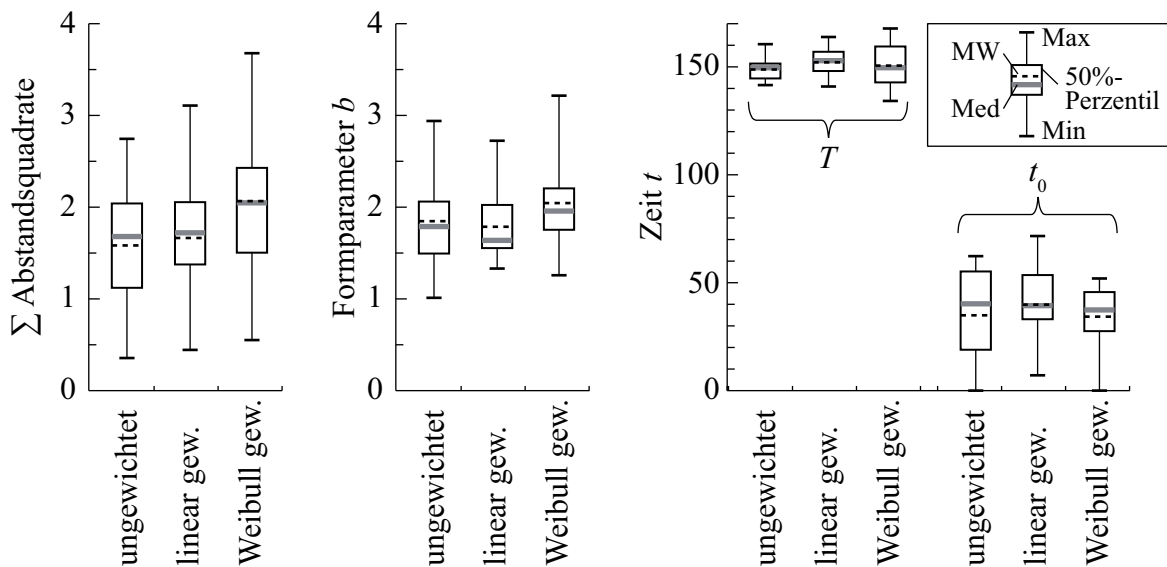


Bild 4.20: Vergleich der ordinatenbasierten Fitnessberechnung ohne Gewichtung, mit linearer Gewichtung und mit Gewichtung in Weibullkoordinaten

Keine Gewichtungsvariante zeigt eine eindeutige Verbesserung aller Parameter. Die lineare Gewichtung bietet eine geringe Verbesserung der Schätzung von b auf Kosten einer etwas schlechteren Schätzung von T . Die Gewichtung in Weibullkoordinaten zeigt einen verringerten Streubereich der ausfallfreien Zeit bei einer ebenfalls höheren Streuung der charakteristischen Lebensdauer.

Insgesamt verbessert die Gewichtung in Weibullkoordinaten die Approximation der Datenpunkte im Bereich geringer Ausfallwahrscheinlichkeiten. Dies ist für die Schätzung von gemischten Verteilungen von entscheidender Bedeutung. Darum kann trotz der Verschlechterung der Anpassung im mittleren Ausfallwahrscheinlichkeitsbereich und der daraus resultierenden Verschlechterung der Summe der Abstandsquadrate diese Gewichtungsvariante als die geeignetste erachtet werden. Deshalb wird im Folgenden die Berechnung der Fitness in Ordinatenrichtung in Verbindung mit einer Gewichtung in Weibullkoordinaten verwendet.

4.4 Optimierung der Parameter des genetischen Algorithmus

Die mathematischen Grundlagen zum Ablauf des genetischen Algorithmus wurden in Kapitel 4.2 für die Kodierung und in Kapitel 4.3 für die Berechnung der Fitness beschrieben. Im nächsten Schritt werden weitere Parameter des genetischen Algorithmus untersucht und die für die Parameterschätzungen von gemischten Verteilungen sinnvollen Größen oder Varianten ausgewählt. Diese weiteren notwendigen Parameter sind in Bild 4.1 dargestellt. Die wichtigsten Stellgrößen stellen die genetischen Operatoren dar, diese werden in den Kapiteln 4.4.1 bis 4.4.3 untersucht. Einen großen Einfluss auf die Güte der Parameterschätzung haben auch die Populations- und Lösungsraumgröße und die Definition der Abbruchbedingung über die Generationenanzahl.

Zur Untersuchung der Auswirkungen der verschiedenen Stellgrößen werden die in Kapitel 4.1 gezeigten Mischverteilungen verwendet. Zum Vergleich der Lösungen dient wie in den vorherigen Kapiteln die Summe der Abstandsquadrate. Es wird die standardbinäre Kodierung und als Fitnessfunktion die Summe der in Weibullkoordinaten gewichteten Abstände in Ordinateurichtung verwendet. Die jeweils verwendeten Initialisierungsparameter sind im Anhang dargestellt. Für jede Variante werden 20 Replikationen durchgeführt.

4.4.1 Rekombination

Der wichtigste genetische Operator ist die Rekombination. Die in dieser Arbeit verwendeten Rekombinationsvarianten wurden bereits in Kapitel 3.1.7 vorgestellt. Diese Varianten werden auf ihren Einfluss auf die Güte der Schätzung von Mischverteilungen hin untersucht. Die zur Schätzung der Verteilungen verwendeten Initialisierungsparameter sind im Anhang in Tabelle A1.3 zu finden.

Die einfachste Variante ist der 1-Punkt-Crossover. Hier wird im Intervall $[1, L_K-1]$ ein Punkt ausgewählt, ab dem die Codesegmente der beiden Eltern eines Individuums vertauscht werden. Aus den beiden Eltern entstehen so zwei Nachkommen. In Bild 4.21 links sind die Ergebnisse für die Parameterschätzungen der Mischverteilungstypen A bis F mit dem 1-Punkt-Crossover dargestellt.

Die Medianwerte der Summe der Abstandsquadrate der Schätzungen liegen alle im Bereich von 1,5 bis 2,5. Die 50 %-Perzentile der Typen A, B, D und F sind im Vergleich zu den Typen C und E sehr klein. Die Typen B und C weisen Ausreißer mit sehr hohen Werten auf. Die Typen C und E sind vermutlich durch ihren s-förmigen Verlauf ohne Stufen und scharfe Knickstellen nur schwer schätzbar. Beim 1-Punkt-Crossover werden sehr große Teile des Codes der Individuen getauscht, wodurch sich die Nachkommen sehr stark von den Eltern unterscheiden, was bei diesen Typen zu sehr schlechten Schätzungen führt. Die Rechenzeit pro Replikation beträgt etwa 5,75 s.

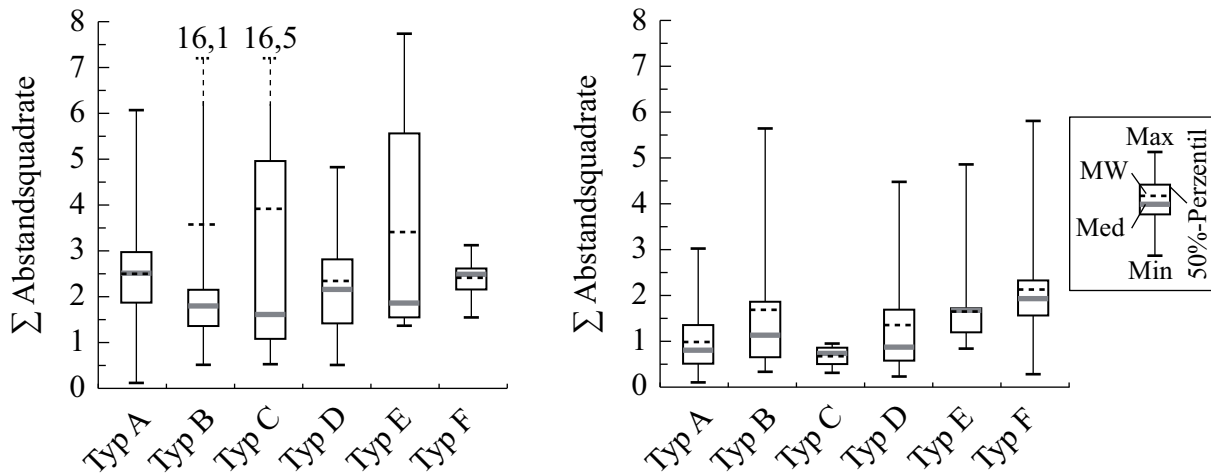


Bild 4.21: Vergleich der Parameterschätzungen mit 1-Punkt-Crossover (links) und N-Punkt-Crossover (rechts)

Die zweite Rekombinationsvariante ist der N-Punkt-Crossover. Dabei wird die Vertauschung der Kodesegmente der Eltern an mehreren zufällig ausgewählten Stellen vorgenommen. Bild 4.21 rechts zeigt die Ergebnisse der Parameterschätzung mit N-Punkt-Crossover. Gegenüber den Schätzungen mit 1-Punkt-Crossover ist eine teilweise deutliche Reduktion der Streuungen und der Medianwerte zu erkennen. Eindeutige Verbesserungen der Schätzungen sind bei den Typen C und E erkennbar. Hier wirken sich die kürzeren Kodesegmente, die vertauscht werden, besonders positiv aus. Die Rechenzeit pro Replikation beträgt etwa 8,39 s. Trotz der längeren Rechenzeit pro Replikation ist der N-Punkt-Crossover dem 1-Punkt-Crossover vorzuziehen.

Die dritte und aufwändigste untersuchte Rekombinationsvariante ist der Uniform-Crossover. Hier werden anhand einer zuvor festgelegten Wahrscheinlichkeit p_{cross} einzelne Bits der Elternindividuen vertauscht. Dadurch kann direkt die Stärke des Einflusses der Veränderung durch die Rekombination definiert werden. Die untersuchten Rekombinations-Wahrscheinlichkeiten wurden in Anlehnung an die von Mühlenbein in [87] und Jansen in [88] vorgeschlagenen Mutationswahrscheinlichkeiten berechnet. Daraus ergeben sich mit der Kodierungslänge von 57 Bits die Wahrscheinlichkeiten $p_{cross} = 0,018$ und $p_{cross} = 0,134$. Zusätzlich wird als weitere Wahrscheinlichkeit $p_{cross} = 0,5$ ausgewählt. Dies stellt gleichzeitig die höchste Rekombinations-Wahrscheinlichkeit dar, da hierbei jeder Nachkomme ungefähr die Hälfte der Kodierung von beiden Elternindividuen erhält.

In Bild 4.22 sind die Ergebnisse für die Parameterschätzungen der Mischverteilungen mit dem Uniform-Crossover für die verschiedenen Rekombinations-Wahrscheinlichkeiten dargestellt. Aufgetragen sind die Box-Plots der Ergebnisse aus 20 Replikationen der Mischverteilungstypen A bis F.

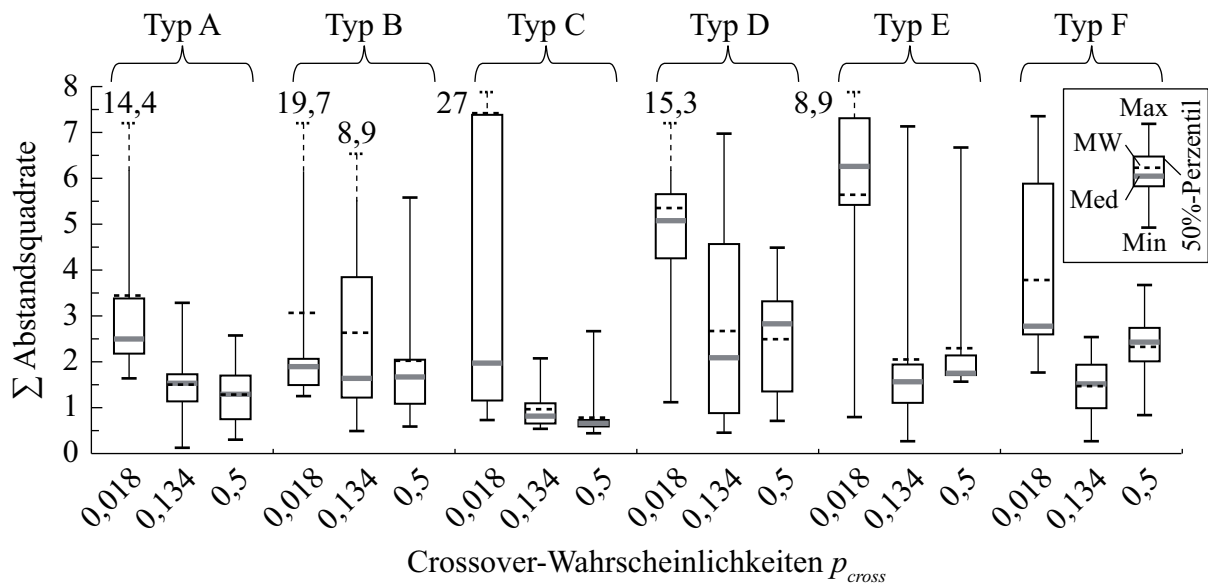


Bild 4.22: Vergleich der Parameterschätzungen mit Uniform-Crossover mit unterschiedlichen Rekombinations-Wahrscheinlichkeiten

Die Ergebnisse der Schätzungen mit der geringsten Rekombinations-Wahrscheinlichkeit liegen in einem ähnlichen Bereich wie das 1-Punkt-Crossover. Auch hier zeigen sich bei den Typen C und E die schlechtesten Schätzungen. Die geringe Wahrscheinlichkeit führt zu wenigen Änderungen bei der Rekombination, wodurch die Nachkommen den Eltern sehr ähnlich sind. Dadurch sind nur geringe Verbesserungen der Fitness möglich. Mit steigender Rekombinations-Wahrscheinlichkeit ist ein Trend zu besseren Medianwerten und geringeren Streuungen zu beobachten. Die besten Medianwerte werden bei den Typen A bis C bei $p_{cross} = 0,5$ erreicht. Am deutlichsten zeigt sich diese Verbesserung bei Typ C. Für die Typen D bis F zeigen sich bei dem mittleren Wert von $p_{cross} = 0,134$ bessere Medianwerte als bei $p_{cross} = 0,5$. Hier scheint die hohe Rekombinations-Wahrscheinlichkeit eine Verbesserung der Nachkommen eher zu behindern. Ein eindeutiges Optimum des Parameters p_{cross} ist bei den vorliegenden Werten für alle Mischverteilungstypen nicht erkennbar.

Die Rechenzeit je Replikation steigt mit zunehmender Rekombinations-Wahrscheinlichkeit von 5,99 s über 6,3 s auf 7,54 s an. Dies ist mit der Anzahl der Vertauschungen begründbar, die mit zunehmender Rekombinations-Wahrscheinlichkeit ansteigt. Zusätzlich ist innerhalb der einzelnen Replikationen eine relativ große Streuung der Rechenzeit erkennbar. Beim 1-Punkt-Crossover liegt diese bei etwa 0,8 % bezogen auf die mittlere Rechenzeit, beim Uniform-Crossover mit $p_{cross} = 0,5$ steigt diese auf 4 % durch die unterschiedlich häufige Anzahl an Vertauschungen an.

Der 1-Punkt-Crossover zeigt mit dem Uniform-Crossover mit geringen Rekombinations-Wahrscheinlichkeiten die schlechtesten Ergebnisse. Mit Erhöhung der Rekombinations-Wahrscheinlichkeit steigt die Güte der Parameterschätzung des Uniform-

Crossover an. Ein Optimum für alle Mischverteilungstypen liegt im Bereich zwischen $p_{cross} = 0,134$ und $p_{cross} = 0,5$.

Der N-Punkt-Crossover zeigt für alle Mischverteilungstypen die besten Parameterschätzungen mit geringen Streubereichen. Trotz der im Vergleich zu den anderen Varianten langen Rechenzeit mit 8,39 s wird im Folgenden wegen der allgemein guten Parameterschätzungen der N-Punkt-Crossover weiterverwendet.

4.4.2 Mutation

Der zweite genetische Operator ist die Mutation, sie kommt nach der Rekombination zum Einsatz. Die Mutation hat zu Beginn, d. h. bei einer geringen Anzahl an durchlaufenen Generationen, im Gegensatz zur Rekombination nur einen geringen Einfluss auf die Diversität der Lösungsvarianten. Mit zunehmender Generationenzahl gleichen sich die Lösungsvarianten immer mehr an, wodurch die Bedeutung der Mutation als zufälliger Modifikator zunimmt.

Die Auftretenshäufigkeit einer Mutation ist über die Mutationswahrscheinlichkeit p_{mut} definiert. Untersucht werden verschiedene Werte für p_{mut} . Als kleinster Wert wurde $p_{mut} = 0,001$ gewählt. Ein weiterer Wert $p_{mut} = 0,018$ berechnet sich nach Jansen [88] mit Gl. (3.3), mit der Kodierungslänge $L_K = 57$. Mit der Empfehlung nach Mühlenbein [87] berechnet sich mit Gl. (3.2) $p_{mut} = 0,134$. Zudem wurden die Werte $p_{mut} = 0,075$ und $p_{mut} = 0,25$ als weitere Werte gewählt. Des Weiteren wurden zwei Varianten eingeführt, bei denen die Mutationswahrscheinlichkeit nach den Kodierungslängen L_i der einzelnen Parameter der Individuen mit Gl. (3.3) und Gl. (3.2) berechnet wird. Mit den im Anhang A1 gegebenen Werten können die Kodierungslängen der einzelnen Parameter berechnet werden. Dies sind $L_b = 13$, $L_T = 26$, $L_{t0} = 1$ und $L_p = 17$. Diese Varianten sind in Bild 4.23 und Bild 4.24 als „Jansen“ und „Mühl“ bezeichnet. Die verwendeten Initialisierungsparameter sind im Anhang in Tabelle A1.4 zu finden.

In Bild 4.23 sind die Ergebnisse der Parameterschätzungen der Mischverteilungstypen A, B und C dargestellt. Die besten Schätzungen der Parameter mit den geringsten Medianwerten der Summe der Abstandsquadrate werden mit der Mutationswahrscheinlichkeit $p_{mut} = 0,018$ nach der Empfehlung von Jansen erreicht. Bei kleineren Mutationswahrscheinlichkeiten erzeugt die Mutation zu wenige neue Individuen, die andere Bereiche des Lösungsraums abdecken. Mit zunehmender Mutationswahrscheinlichkeit werden die Eigenschaften guter Individuen durch die Mutation wieder verschlechtert, was insgesamt zu schlechteren Lösungsvarianten führt. Die Varianten „Jansen“ und „Mühl“ mit den aus den einzelnen Kodierungslängen berechneten Mutationswahrscheinlichkeiten zeigen bis auf Typ A vergleichbare Ergebnisse wie die Variante mit $p_{mut} = 0,018$. Die Variante nach Mühlenbein liefert bei Anwendung auf die einzelnen

Kodierungslängen eine bessere Schätzung als bei Anwendung auf die gesamte Kodierungslänge L_K , ist aber etwas schlechter als die Varianten nach Jansen. Die Rechenzeit bewegt sich im Mittel bei 169 s für 20 Replikationen für alle Varianten annähernd gleich.

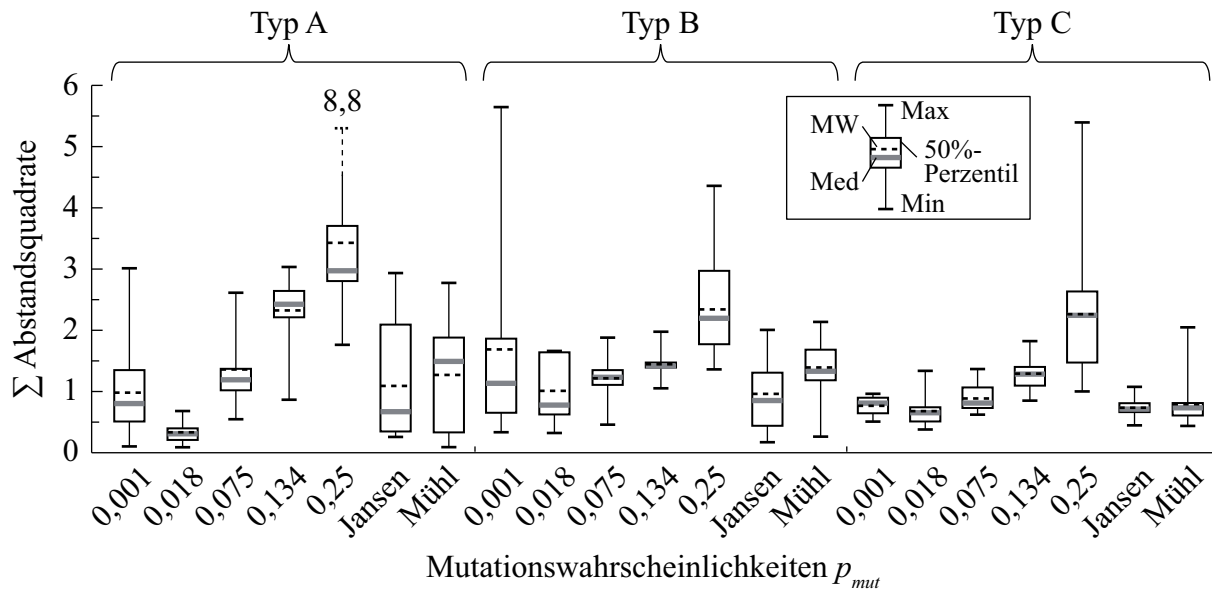


Bild 4.23: Vergleich der Parameterschätzungen mit verschiedenen Mutationswahrscheinlichkeiten für die Mischverteilungstypen A bis C

Die Ergebnisse der Parameterschätzung für die Mischverteilungstypen D, E und F sind in Bild 4.24 dargestellt.

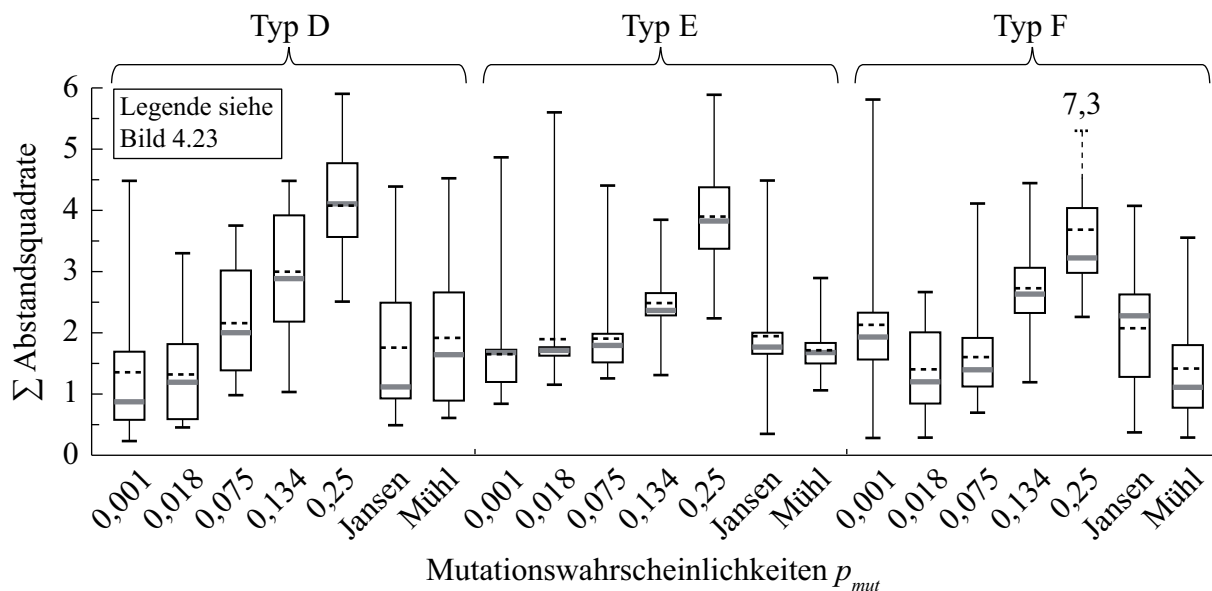


Bild 4.24: Vergleich der Parameterschätzungen mit verschiedenen Mutationswahrscheinlichkeiten für die Mischverteilungstypen D bis F

Für die Typen D bis F ergibt sich ein ähnliches Ergebnis wie in Bild 4.23 für die Typen A bis C. Hohe Mutationswahrscheinlichkeiten führen zu schlechteren

Schätzungen. Für den Typ D zeigt sich abweichend jedoch ein besserer Medianwert für $p_{mut} = 0,001$ als für $p_{mut} = 0,018$.

Es zeigt sich, dass für die Schätzung der Mischverteilungen ein Optimum für die Mutationswahrscheinlichkeit im Bereich von $p_{mut} = 0,018$ liegt. Geringere Werte bewirken eine zu geringe Abdeckung des Lösungsraumes, größere eine Verschlechterung der schon gefundenen guten Lösungsvarianten. Die Varianten „Mühl“ und „Jansen“ bewirken keine Verbesserung der Parameterschätzung. Deshalb wird für weitere Untersuchungen die von Jansen vorgeschlagene Variante mit einer Mutationswahrscheinlichkeit für die gesamte Kodierung eines Individuums genutzt.

4.4.3 Selektion

Die Selektionsmethode der Individuen innerhalb einer Generation hat den größten Einfluss auf die Konvergenz der Lösungen eines genetischen Algorithmus. Durch ungünstige Wahl einer Selektionsstrategie kann es zu einer langsamen Konvergenz oder sogar zu einer Divergenz der Lösungen kommen. In Bild 4.2 sind die Paarungs-, Populations- und Umweltselektion im Ablauf des genetischen Algorithmus dargestellt.

Für die Paarungs- und Umweltselektion sind prinzipiell alle in Kapitel 3.1.8 näher erläuterten Selektionsvarianten denkbar. Die Populationsselektion unterscheidet sich jedoch, da mit ihr festgelegt wird, ob und welche Anteile an Eltern und Nachkommen für die Umweltselektion zur Verfügung stehen.

An die Selektionen werden zwei gegensätzliche Anforderungen gestellt:

- Zulassen einer möglichst großen Vielfalt an Lösungsmöglichkeiten
- Bevorzugen der besten Individuen

Mit der ersten Anforderung wird verhindert, dass sich gute Individuen zu stark auf die Population auswirken, wodurch die Lösung schnell zu einem lokalen Optimum konvergieren kann und der gesamte Lösungsraum schlecht abgedeckt wird. Nachteilig kann sich dies auswirken, wenn gute Individuen nur wenig berücksichtigt werden oder ganz aus dem Zyklus ausscheiden. Die zweite Anforderung ist der ersten entgegengesetzt und bewirkt im positiven Fall eine schnelle Konvergenz und im negativen Fall eine Konvergenz zu einem gegenüber dem globalen Optimum schlechten lokalen Optimum. Es muss ein Mittelweg zwischen den beiden Anforderungen gefunden werden, mit dem eine hinreichend schnelle Konvergenz und eine möglichst gute Abdeckung des Lösungsraums erreicht wird.

Mit Hilfe dieser beiden Anforderungen können die 16 möglichen Kombinationen der Paarungs- und Umweltselektion eingeschränkt werden. Eine der beiden Selektionen sollte eine Selektionsvariante sein, die eine große Vielfalt an Lösungen zulässt, die

zweite eine Selektionsvariante, die die besten Individuen bevorzugt. Damit sind beide Anforderungen innerhalb eines Durchlaufs erfüllt.

Eine andere Möglichkeit besteht darin, für beide Selektionen Selektionsvarianten zu wählen, die sowohl zufällige als auch fitnessabhängige Komponenten beinhalten, beispielsweise die rangproportionale Selektion. Die Fitnessselektion und die fitnessproportionale Selektion sind wegen der sehr starken Gewichtung des Fitnesswertes weniger geeignet; ebenso die stochastische Selektion wegen ihrer rein zufälligen Auswahl der Individuen. Die untersuchten Kombinationen der Selektionsvarianten für die Paarungs- und Umweltselektion sind in Tabelle 4.3 links aufgeführt.

Tabelle 4.3: Varianten für Paarungs- und Umweltselektion sowie Populationsselektion

Varianten	Paarungsselektion	Umweltselektion	Populationsselektion
Var_1	stochastisch	Fitnessselektion	nur Nachkommen
Var_2		fitnessproportional	alle Eltern + Nachkommen
Var_3		rangproportional	50 % Eltern + 50 % Nachkommen
Var_4	rangproportional	rangproportional	20 % Eltern + 80 % Nachkommen

Bei der Auswahl der Populationsselektion ist entscheidend, ob nur Nachkommen oder Eltern und Nachkommen für die Umweltselektion vorhanden sein sollen. Zusätzlich kann über die Anteile der Eltern an der Population für die nächste Generation bestimmt werden, wie viele neue Individuen in der nachfolgenden Population vorhanden sind. Die verschiedenen untersuchten Varianten der Populationsselektion sind in Tabelle 4.3 rechts dargestellt.

Die Auswirkungen der verschiedenen Möglichkeiten der Populationsselektion auf den Fitnesswert und die Veränderung der Fitness über die Generationen sind beispielhaft für die Mischverteilungstypen A und F in Bild 4.25 dargestellt.

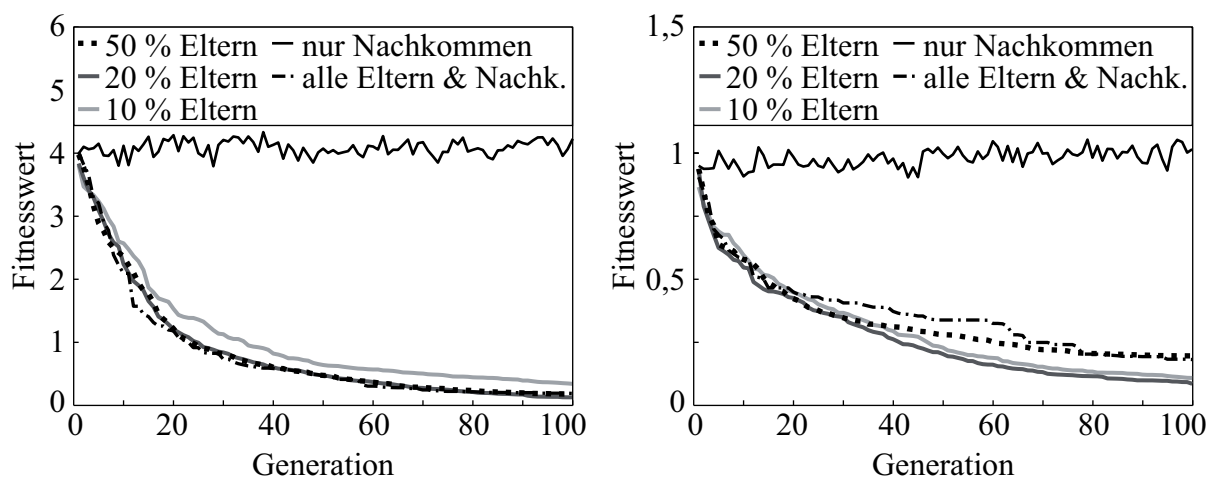


Bild 4.25: Vergleich der Verläufe des Fitnesswertes für die Mischverteilungstypen A (links) und F (rechts) bei verschiedenen Anteilen an Elternindividuen

Die verwendeten Initialisierungsparameter sind im Anhang in Tabelle A1.5 zu finden. Als Paarungsselektion wurde die stochastische Selektion, als Umweltselektion die Fitnessselektion gewählt. Dargestellt ist der Median des Fitnesswertes aller 20 Replikationen in der jeweiligen Generation für die untersuchten Varianten. Bei der Variante, die nur Nachkommen in der nachfolgenden Generation zulässt, ist keine Konvergenz zu einem Grenzwert zu erkennen. Obwohl nur die besten Individuen der Nachkommen in die neue Generation übernommen werden, ist allein durch die genetischen Operatoren keine Verbesserung der Gesamtpopulation garantiert. Ein hoher Elternanteil in der Population erschwert das Finden neuer Optima, da zu wenige neue Individuen weitergegeben werden. Die Variante, bei der alle Eltern und Nachkommen für die Umweltselektion zur Verfügung stehen, verhält sich, durch den hohen Elternanteil, wie die Variante mit festgelegtem 50 % Elternanteil an der Population. Ein zu geringer Elternanteil führt zu einem Verlust an guten Elternindividuen für die folgende Generation.

Bei allen untersuchten Mischverteilungen konnte mit einem Anteil von 20 % Eltern bei der Populationsselektion das beste Ergebnis erzielt werden.

Bei der Untersuchung der verschiedenen Varianten der Paarungs- und Umweltselektion wird die Populationsselektion mit einer Population aus 20 % Elternindividuen und 80 % Nachkommenindividuen gewählt. Die untersuchten Varianten sind in Tabelle 4.3 mit Var_1 bis Var_4 bezeichnet. In Bild 4.26 sind die Ergebnisse der untersuchten Varianten dargestellt.

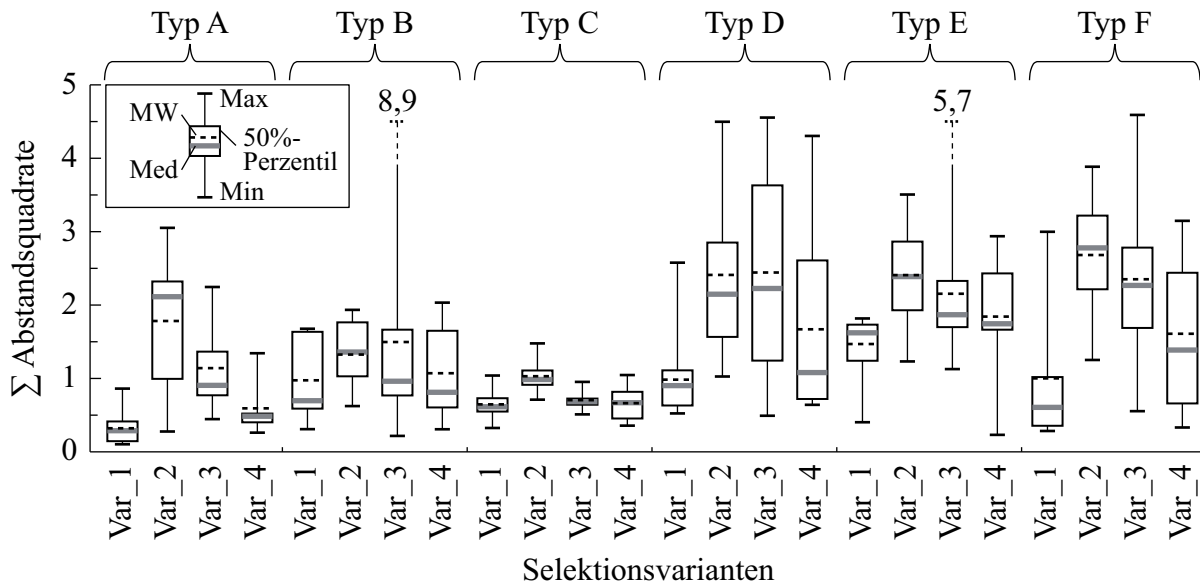


Bild 4.26: Vergleich der Parameterschätzungen der Mischverteilungstypen A bis F mit verschiedenen Paarungs- und Umweltselektionsvarianten

Für alle Mischverteilungstypen zeigt Var_1 mit der stochastischen Selektion als Paarungsselektion und der Fitnessselektion als Umweltselektion die besten Ergebnisse. Bei den beiden Varianten Var_2 und Var_3 mit stochastischer Paarungsselektion und

fitness- bzw. rangproportionaler Selektion als Umweltselektion sind die Medianwerte und teilweise auch die Streubereiche deutlich größer. Dies deutet darauf hin, dass gute Individuen zu häufig in der Population vorhanden sind und dadurch die Vielfalt der Individuen zu stark eingeschränkt wird. Die zweitbeste Variante ist Var_4, bei der zweimal eine fitnessabhängige Selektionsvariante, die rangproportionale Selektion, verwendet wird. Anhand von Bild 4.27 ist erkennbar, aus welchem Grund diese Variante schlechtere Ergebnisse liefert als Var_1. Aufgetragen ist der Median des Fitnesswerts über den Generationen.

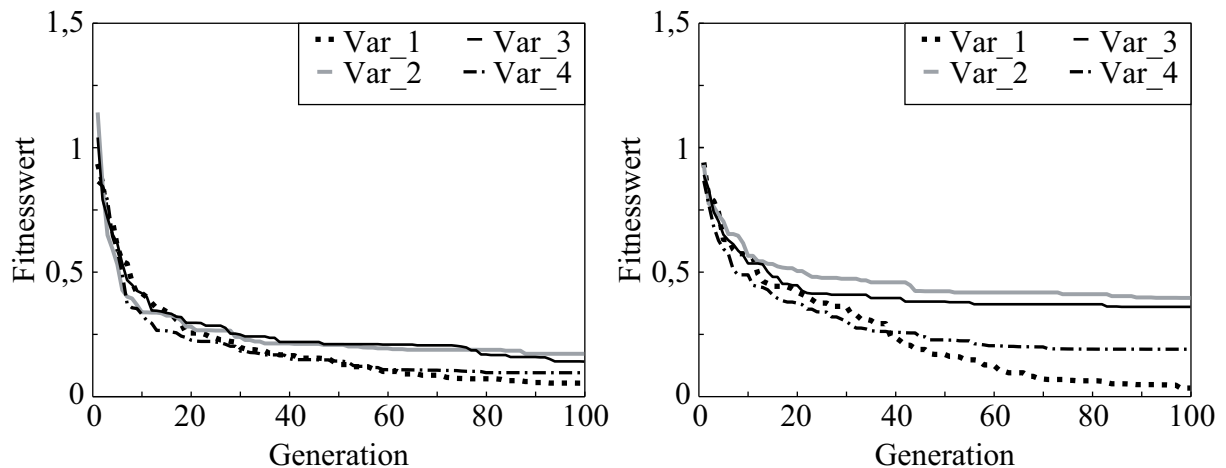


Bild 4.27: Vergleich der Verläufe des Fitnesswertes für die Mischverteilungstypen *D* (links) und *F* (rechts) bei verschiedenen Paarungs- und Umweltselektionsvarianten

Während bei Var_1 die Fitness kontinuierlich auch bei hoher Generationenanzahl noch abnimmt, ist bei Var_4 ungefähr ab der 60. Generation keine Verbesserung der Fitness mehr erkennbar. Hier wird die Vielfalt in der Population zu stark durch ähnliche Individuen mit geringer Fitness eingeschränkt. Dies ist auch bei Var_2 und Var_3 zu sehen, bei denen die Gewichtung guter Individuen noch stärker ist.

Die besten Selektionsvarianten zur Parameterbestimmung von Mischverteilungen können eindeutig definiert werden:

- Eine stochastische Selektion als Paarungsselektion, um die Vielfalt innerhalb einer Population zu vergrößern und dadurch bessere Individuen innerhalb des Lösungsraums zu finden.
- Eine Populationsselektion mit einem Anteil von 20 % Elternindividuen, um aus der Vielzahl an Nachkommen bessere Lösungsvarianten zu gewinnen und um durch die Übernahme von Eltern die Güte der Population beizubehalten.
- Eine Fitnessselektion als Umweltselektion, um aus den Eltern und Nachkommen nur die besten Individuen auszuwählen und so eine Konvergenz zu einem Ergebnis zu ermöglichen.

4.4.4 Einfluss der Generationenanzahl, Populations- und Suchraumgröße

Einen weiteren großen Einfluss auf die Güte der Schätzung haben die für die Initialisierung des genetischen Algorithmus erforderliche Populationsgröße und die Definition der Abbruchbedingung, siehe Bild 4.1. Die Abbruchbedingung wird über die Anzahl der durchlaufenen Generationen definiert. Eine weitere Einflussgröße, die sich aus der Initialisierung der Problemstellung ergibt, ist die Größe des Lösungsraums und die erforderliche Genauigkeit der Parameter.

Anhand des Mischverteilungstyps E werden beispielhaft die oben genannten Einflüsse untersucht. Die dazu verwendeten Parameter des genetischen Algorithmus sind im Anhang in Tabelle A1.6 zu finden. Bild 4.28 zeigt den Einfluss der Populationsgröße auf die Parameterschätzung.

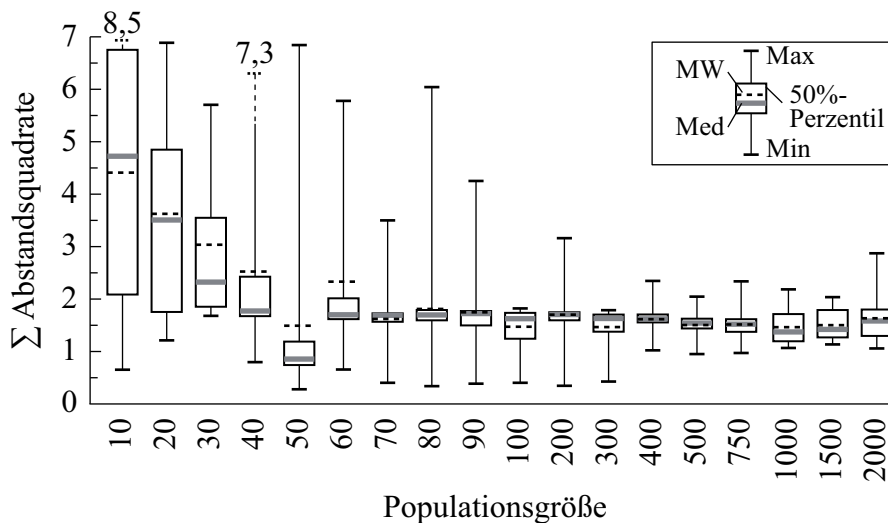


Bild 4.28: Einfluss der Populationsgröße auf die Parameterschätzung des Mischverteilungstyps E

Die Abbruchbedingung, der Lösungsraum und die Genauigkeit der Parameter sind identisch mit den in Kapitel 4.4.3 verwendeten. Bei einer geringen Populationsgröße zeigt sich eine große Streuung und im Mittel ein schlechter Wert für die Summe der Abstandsquadrate. Mit zunehmender Größe der Population verbessern sich der Medianwert und die Streuung und erreichen bei jeweils 70 Eltern und Nachkommen ein Minimum des 50 %-Perzentils. Darüber hinaus findet keine Verbesserung mehr statt. Im Bereich einer Populationsgröße von 50 bis 300 sind im Idealfall bessere Lösungen möglich, die bei größeren Populationen in der Selektion nicht mehr berücksichtigt werden.

Ein ähnliches Verhalten zeigt sich bei der Betrachtung des Einflusses der Anzahl der Generationen, siehe Bild 4.29. Die verwendete Populationsgröße liegt bei 100, der Lösungsraum und die Genauigkeit der Parameter sind wieder mit den in Kapitel 4.4.3 verwendeten identisch.

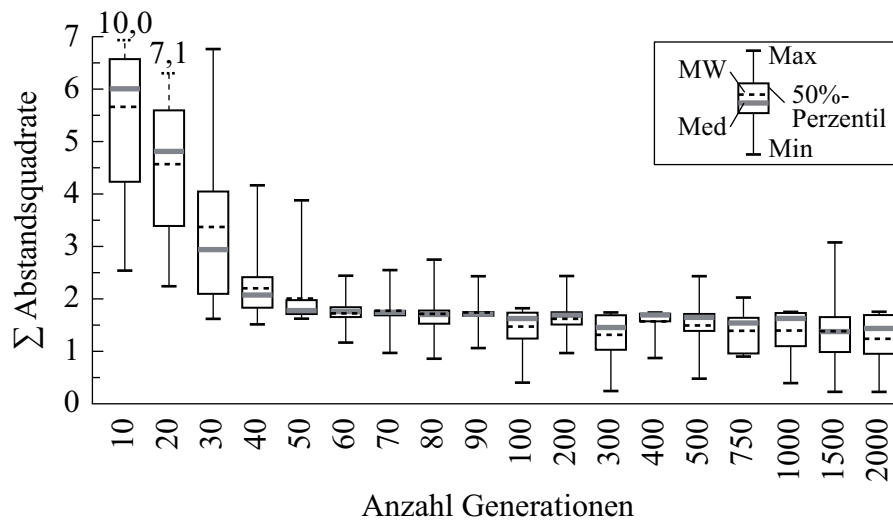


Bild 4.29: Einfluss der Generationenanzahl auf die Parameterschätzung des Mischverteilungstyps E

Für den Mischverteilungstyp E strebt die Summe der Abstandsquadrate von einem hohen Niveau bei geringer Generationenanzahl zu einem Optimum, das bei einer Generationenanzahl von 60 bereits erreicht ist. Eine Erhöhung der Generationenanzahl bewirkt nahezu keine weitere Verbesserung. Das Verharren des Medianwertes sowohl in Bild 4.28 als auch in Bild 4.29 bei etwa 1,5 deutet auf ein starkes lokales Optimum der Fitnessfunktion bei Mischverteilungstyp E hin.

Als letzte Einflussgrößen werden die Lösungsraumgröße und die Genauigkeit der Parameter untersucht. Die Größe des Lösungsraumes wird in drei Stufen von L1 bis L3 variiert, ebenso die Genauigkeit der Parameter mit den Stufen G1 bis G3. Die Werte können Tabelle 4.4 entnommen werden. Weiterhin werden die Anzahl an Generationen mit 100 und die Populationsgröße mit 100 gewählt.

Tabelle 4.4: Varianten des Lösungsraums und der Genauigkeit der Parameter

	Lösungsraumgröße			Genauigkeit der Parameter (m_b, m_r, m_o, m_p)		
	L1	L2	L3	G1	G2	G3
$\begin{pmatrix} b_u & b_o \\ T_u & T_o \\ t_{0u} & t_{0o} \\ p_u & p_o \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0,5 & 4 \\ 0,5 & 100 \\ 0 & 0 \\ 0,2 & 0,8 \end{pmatrix}$	$\begin{pmatrix} 0,7 & 3,5 \\ 0,7 & 30 \\ 0 & 0 \\ 0,25 & 0,75 \end{pmatrix}$	(3, 5, 0, 5)	(2, 3, 0, 3)	(1, 1, 0, 2)

Die Größe des Lösungsraumes für die Parameter verringert sich von L1 nach L3. Ebenso verringert sich die Genauigkeit der Parameter von G1 nach G3. Die Ergebnisse der Parameterschätzung sind in Bild 4.30 dargestellt.

Der Einfluss der Größe des Lösungsraumes ist an der Verringerung des Medianwertes von L1 nach L3 erkennbar; je kleiner der Lösungsraum ist, umso geringer ist die

Abweichung zu den Datenpunkten. Die Genauigkeit hat einen Einfluss auf die Streuung der Ergebnisse. Eine hohe Genauigkeit (G1) mit vielen Dezimalstellen bewirkt eine Streuung aufgrund der Vielzahl an sehr ähnlichen Lösungsvarianten. Eine geringe Genauigkeit (G3) führt zu einer Vergrößerung des Streubereichs, bedingt durch eine zu geringe Anpassungsfähigkeit an das Optimum. Die kleinsten 50 %-Perzentile werden mit der mittleren Genauigkeit erreicht, die eine ausreichende Anpassungsfähigkeit an das Optimum ermöglicht.

Erkennbar ist auch, dass das starke lokale Optimum bei einer Summe der Abstandsquadrate von 1,5 durch die Verkleinerung des Lösungsraumes aus den möglichen Lösungsvarianten herausfällt.

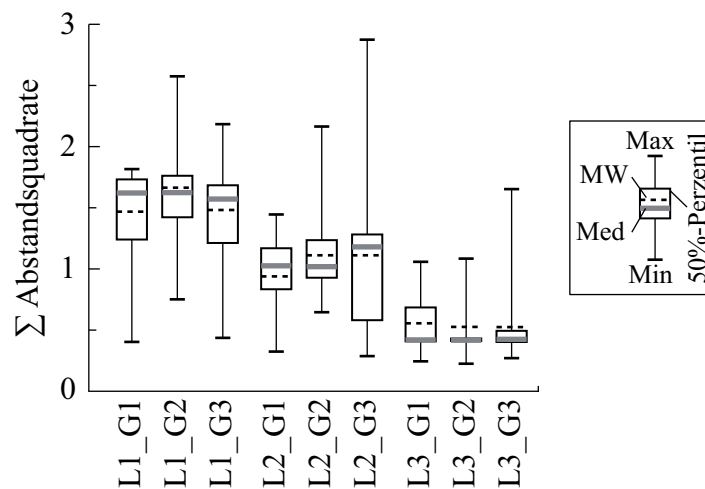


Bild 4.30: Einfluss der Größe des Lösungsraums und der Genauigkeit der Parameter auf die Parameterschätzung des Mischverteilungstyps E

Über die Genauigkeit der Parameter und die Lösungsraumgröße lässt sich die Anzahl der möglichen Lösungsvarianten der jeweiligen Problemstellung berechnen. Aus den Intervallen der Parameter $[x_u, x_o]$, der gewählten Genauigkeit des Parameters m_x , der Anzahl der Parameter u und der Anzahl der Unterverteilungen N ergibt sich mit Gl. (4.26) die Menge der Lösungsvarianten M .

$$M = N \sum_{i=1}^u (x_{o,i} - x_{u,i}) 10^{m_{x,i}} \quad (4.26)$$

Über die Anzahl der Generationen G und die Größe der Population ζ kann mit Gl. (4.27) aus der Menge der Lösungsvarianten M die theoretische Lösungsdichte ρ_L berechnet werden.

$$\rho_L = \frac{G\zeta}{M} \quad (4.27)$$

Die theoretische Lösungsdichte gibt an, wie viele der möglichen Lösungsvarianten während einer Replikation des genetischen Algorithmus vorkommen können, wenn bei

jeder Generation vollkommen neue Lösungsvarianten erzeugt würden. Die eigentliche Lösungsdichte liegt weit unterhalb der theoretischen Lösungsdichte. Es werden in späteren Generationen sehr viele gleiche Lösungen erzeugt, da die Elterngenerationen sich nicht mehr stark unterscheiden. Eine Erhöhung der Generationenanzahl führt deshalb nicht notwendigerweise zu einer Verbesserung der Güte der Schätzung. Der Einfluss einer vergrößerten Populationsanzahl bietet demgegenüber den Vorteil einer größeren Variantenvielfalt. Die Vielfalt der Lösungsvarianten wird durch die genetischen Operatoren erzeugt, wodurch bei einer größeren Population die Wahrscheinlichkeit für neue Individuen steigt.

Die beispielhafte Untersuchung der Populationsgröße, der Generationenzahl, der Lösungsraumgröße und der Genauigkeit der Parameter zeigt die Abhängigkeiten der Güte der Parameterschätzung von diesen Algorithmenparametern auf. Eine Anpassung an die jeweilige Problemstellung führt zu besseren Ergebnissen bei geringeren Rechenzeiten.

4.5 Zusammenfassung

Die Untersuchungen der verschiedenen Parameter des Algorithmus zeigen eine starke Abhängigkeit der Ergebnisgüte von den Einstellparametern. Die gefundenen Parameter stellen für die vorliegende Problemstellung ein Optimum für die Mehrheit der Mischverteilungen dar. Durch die große Vielfalt an möglichen gemischten Verteilungen ist eine Optimierung auf einen bestimmten Mischverteilungstyp nicht sinnvoll. Das Verfahren soll allgemein für verschiedene Arten von gemischten Verteilungen verwendbar sein und zu guten Ergebnissen führen, ohne zuvor eine Vielzahl an verschiedenen Parameterkombinationen des Algorithmus auf die Problemstellung hin optimieren zu müssen. Die für die Anwendung auf Mischverteilungen gefundenen Einstellparameter des genetischen Algorithmus sind:

- Kodierung in standardbinärer Darstellung
- Berechnung der Fitness in Ordinateurichtung mit weibullskaliertem Gewichtung
- N-Punkt-Crossover als Rekombinationsvariante
- Mutationswahrscheinlichkeit berechnet nach Jansen [88]
- Selektionstypen:
 - Stochastische Selektion als Paarungsselektion
 - Populationsselektion mit 20 % Elternindividuen und 80 % Nachkommen
 - Fitnessselektion als Umweltselektion
- Populationsgrößen von 100 Eltern- und Nachkommenindividuen ausreichend
- Generationenanzahl von etwa 100 ausreichend
- Lösungsraum und Genauigkeit der Verteilungsparameter so klein wie möglich

Ein Punkt, der in den Untersuchungen in Kapitel 4.2 bis Kapitel 4.4 nicht näher betrachtet wurde, sind die Wechselwirkungen zwischen den einzelnen Algorithmen- und Problemstellungsparametern. Diese Wechselwirkungen können durch eine andere Kombination der Parameter und Einstellgrößen des genetischen Algorithmus zu besseren Konvergenzeigenschaften und Lösungen der Parameterschätzungen führen.

Starke Wechselwirkungen bestehen zwischen:

- Rekombinationsvariante und Mutationswahrscheinlichkeit,
- Rekombination und Mutation mit der Selektionsvariante und
- der Größe des Lösungsraumes und der Genauigkeit der Ergebnisse.

Die Wechselwirkungen haben einen großen Einfluss auf die Abdeckung des Lösungsraumes, welche direkt die Ergebnislösung beeinflusst. Hier gilt es, wie bei der Selektion eine Kombination zu erreichen, die dem Widerspruch zwischen der bestmöglichen Abdeckung des Lösungsraumes mit Individuen und der Konvergenz des genetischen Algorithmus gegen ein Optimum bestmöglich genügt.

5 Gemischte Verteilungen zur Evaluation

Zur Evaluation der Ergebnisgüte und der Einsatzgrenzen des genetischen Algorithmus werden in dieser Arbeit verschiedene repräsentative Datensätze verwendet. Da es sich beim genetischen Algorithmus um eine stochastische Methode handelt, ist die Güte der Parameterschätzung sehr stark von den zu schätzenden Daten abhängig. Dadurch ist eine pauschale Aussage über die Anwendbarkeit der Methode für verschiedene Verteilungen nicht möglich. Deshalb wird das Verhalten des genetischen Algorithmus mit einer möglichst großen Vielfalt an gemischten Verteilungen untersucht. Die in diesem Abschnitt vorgestellten Verteilungen stellen einen repräsentativen Querschnitt der möglichen gemischten Verteilungen dar. Die hier vorgestellten Datensätze wurden zum Teil bereits zur Evaluation und Optimierung der Parameter des genetischen Algorithmus eingesetzt, siehe Kapitel 4.1.

Aus den gegebenen gemischten Verteilungen werden mittels der Medianränge über die Inverse der Verteilungsfunktion Ausfalldaten erzeugt. Diese liegen genau auf der Verteilungsfunktion, siehe Bild 5.1 links. Zusätzlich werden mittels der Monte-Carlo Methode Ausfalldaten erzeugt, die zufällig um die gegebene Verteilungsfunktion streuen, siehe Bild 5.1 rechts. Dies dient zur Untersuchung der Güte der Parameterschätzungen bei einer stochastischen Streuung.

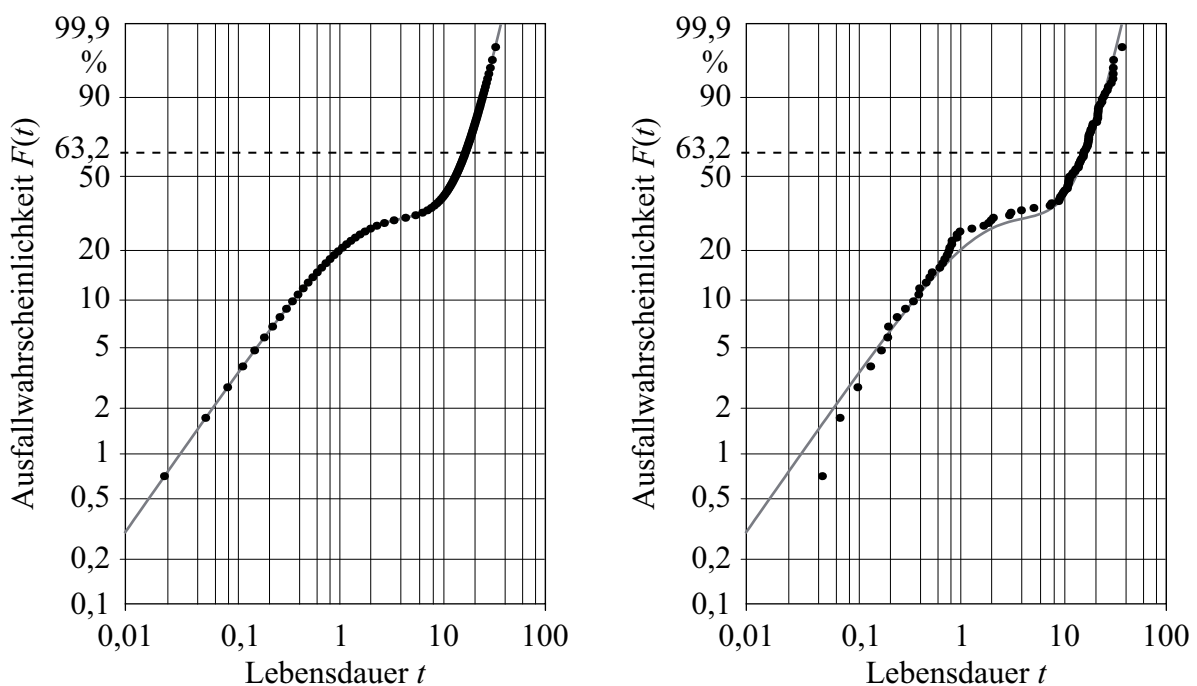


Bild 5.1: Generierte Daten aus Mischverteilungen über Medianränge (links) und Monte-Carlo Methode (rechts)

Durch die Verwendung von synthetisch erzeugten Datensätzen sind die zu schätzenden Parameter der Verteilungen bekannt, die Güte der Schätzung kann damit direkt beurteilt werden. Die erzeugten Daten dienen zusätzlich als Eingangswerte für analytische Vergleichsverfahren.

Im ersten Teil werden die verwendeten Mischverteilungen nach Jiang und Kececioglu [9], die in Kapitel 4.1 eingeführt wurden, vorgestellt. Daraus werden zusätzlich verschiedene Mischverteilungen abgeleitet. Dies sind:

- Mischverteilungen erweitert um eine ausfallfreie Zeit t_0 ,
- Mischverteilungen mit drei Unterverteilungen und
- Mischverteilungen mit vier Unterverteilungen.

Im zweiten Teil werden verschiedene boolesche Serien- und Parallelsysteme vorgestellt. Dabei ist wie auch im ersten Teil die Untersuchung der Separierbarkeit der Verteilungen notwendig. Aus den Überlegungen zur Separierbarkeit von booleschen Systemen werden Verteilungen von Serien- und Parallelsystemen definiert. Diese Serien- und Parallelsysteme werden zusätzlich um den Parameter t_0 erweitert.

Zuletzt wird auf die Schätzung zensierter Datensätze eingegangen und eine mögliche Vorgehensweise aufgezeigt.

Der Stichprobenumfang der verwendeten Datensätze wird zwischen 20, 50 und 100 variiert. Damit wird den üblicherweise geringen Stichprobenumfängen in der Zuverlässigkeitstechnik Rechnung getragen. Dies ist vor allem für die Maximum-Likelihood Methode entscheidend, deren Parameterschätzung am sensibelsten auf den Stichprobenumfang reagiert [16], [121].

5.1 Mischverteilungen

Zur Evaluation der Anwendbarkeit von genetischen Algorithmen und neuronalen Netzen auf Mischverteilungen werden die in diesem Abschnitt vorgestellten Mischverteilungstypen verwendet. Die Mischverteilungen wurden teilweise bereits in Kapitel 4 zur Evaluation der Parameter des genetischen Algorithmus eingesetzt.

5.1.1 Mischverteilungen nach Jiang und Kececioglu

Jiang beschreibt in [41] den Verlauf verschiedener Verteilungsfunktionen von Mischverteilungen mit zwei Unterverteilungen in Abhängigkeit vom Überlappungsbereich ihrer Dichtefunktionen. Die Größe des Überlappungsbereichs ist ein Maß für die Separierbarkeit zweier Unterverteilungen in einer Mischverteilung. In Kapitel 2.2.2 ist dargestellt, wie sich durch die Wahl des Formparameters b und der charakteristischen Lebensdauer T einfach separierbare Mischverteilungen erzeugen lassen.

Die Grundlage zur Betrachtung der Separierbarkeit von Mischverteilungen aus zwei Unterverteilungen liefert Gl. (5.1).

$$F_M(t) = p_1(1 - e^{-\left(\frac{t}{T_1}\right)^{b_1}}) + (1 - p_1)(1 - e^{-\left(\frac{t}{T_2}\right)^{b_2}}) \quad (5.1)$$

Durch die Substitutionen

$$b_v = \frac{b_2}{b_1} \quad (5.2)$$

$$T_v = \left(\frac{T_2}{T_1}\right)^{b_1} \quad (5.3)$$

kann die Separierbarkeit vereinfacht dargestellt werden. Mit dieser Vereinfachung untersuchten Jiang und Kececioglu verschiedene Parameterkombinationen und konnten sechs verschiedene Typen von Mischverteilungen identifizieren. In Bild 5.2 sind die Auftretensbereiche der verschiedenen Mischverteilungstypen in Abhängigkeit von p_1 und der in Gl. (5.2) und Gl. (5.3) definierten Verhältnisse von b und T dargestellt.

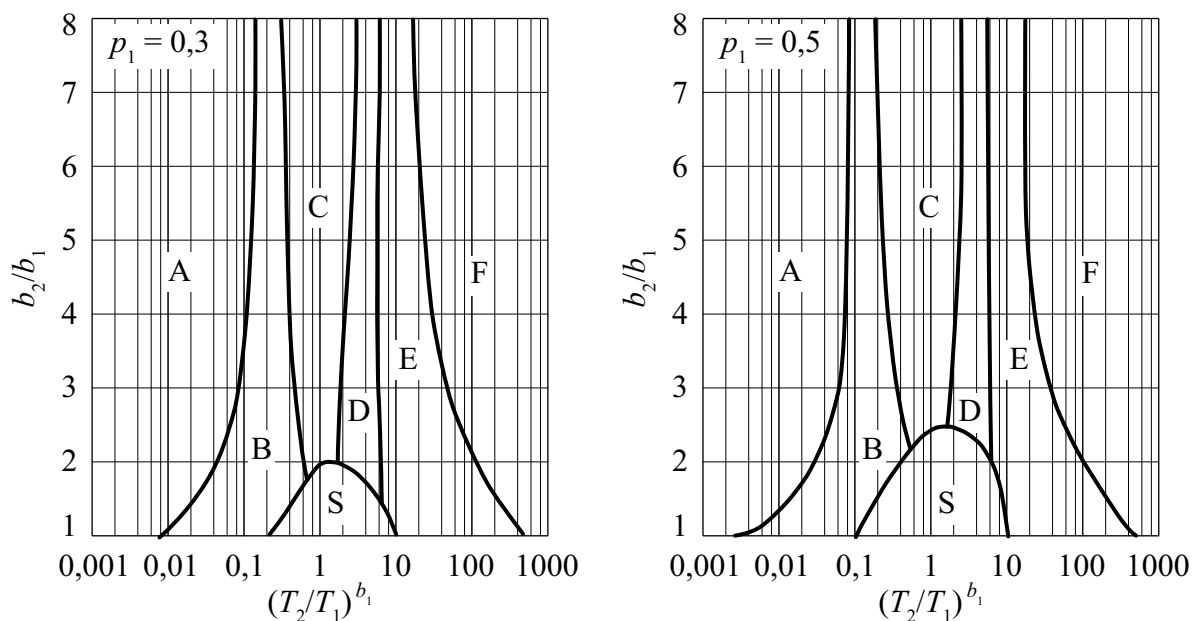


Bild 5.2: Mischverteilungstypen mit zwei Unterverteilungen mit Gewichtungsfaktor $p_1 = 0,3$ (links) und $p_1 = 0,5$ (rechts) nach Jiang [41]

In Bild 5.2 sind die jeweiligen Bereiche den Mischverteilungstypen A bis F zugeteilt. Der Bereich S kennzeichnet den Bereich, in dem die beiden Unterverteilungen einander so ähnlich sind, dass sie nicht mehr separiert werden können und eine Gerade im Weibullpapier ergeben. Als Sonderfall fällt eine Einzelverteilung mit $b_v = 1$ und $T_v = 1$ in den Bereich S. Mit der Angleichung der beiden Gewichtungsfaktoren bei $p_1 = 0,5$, siehe Bild 5.2 (rechts), vergrößert sich der Bereich S der nicht mehr separierbaren

Mischverteilungen stark, wohingegen die Bereiche A bis F nahezu identisch sind. Die Bezeichnungen dieser Bereiche entsprechen den Bezeichnungen der bereits in Kapitel 4.1 vorgestellten Mischverteilungen.

5.1.2 Abgeleitete Mischverteilungen mit mehreren Unterverteilungen

Zur Evaluation der Parameterschätzungen von Mischverteilungen mit mehr als zwei Unterverteilungen werden aus den repräsentativen Mischverteilungstypen von Jiang und Kececioglu weitere Mischverteilungen generiert. Dazu werden die Typen A bis F zu neuen Mischverteilungen mit weiteren Unterverteilungen kombiniert.

Die Parameter der abgeleiteten Mischverteilungen mit drei Unterverteilungen sind in Tabelle 5.1 aufgeführt. Die Mischverteilungen wurden so gewählt, dass sie sich in ihrer Separierbarkeit unterscheiden. Die Unterverteilungen sind über die Gewichtungsfaktoren p_i in etwa gleich gewichtet. Daraus resultieren gleiche Anteile der simulierten Ausfalldaten. Die Namensgebung richtet sich nach den jeweils verwendeten Mischverteilungstypen.

Tabelle 5.1: Parameter der abgeleiteten Mischverteilungen mit drei Unterverteilungen

	b_1	T_1	p_1	b_2	T_2	p_2	b_3	T_3	p_3
Typ AB	1	1	0,3	3	0,01	0,3	3	0,2	0,4
Typ BC	1	1	0,3	3	0,2	0,3	3	1	0,4
Typ EF	1	1	0,3	3	20	0,3	3	100	0,4
Typ AF	1	1	0,3	3	100	0,3	3	0,01	0,4

In Bild 5.3 sind die verwendeten Mischverteilungen mit drei Unterverteilungen abgebildet. Die Merkmale der abgeleiteten Mischverteilungstypen sind:

- Typ AB ist eine Kombination der Typen A und B. Der Verlauf enthält einen Sattelpunkt und eine Knickstelle, wodurch die Unterverteilungen gut den einzelnen Bereichen der Mischverteilung zugeordnet werden können.
- Typ BC enthält die Unterverteilungen der Typen B und C. Durch die gleichen charakteristischen Lebensdauern bei T_1 und T_3 ist das Auftreten von Unterverteilung eins nur an der Krümmung der Mischverteilung im Bereich $F(t) < 10\%$ erkennbar. Dadurch sind die Parameter von Typ BC nur schwer schätzbar.
- Typ EF entsteht durch die Kombination der Typen E und F. Im Verlauf der Kurve sind zwei Stufen zu erkennen, die nicht so stark ausgeprägt sind wie bei Typ AB.
- Typ AF besteht aus den Verteilungen der Typen A und F. In diesem Kurvenverlauf sind zwei Sattelpunkte erkennbar, wodurch die Bereiche der Unterverteilungen klar getrennt sind.

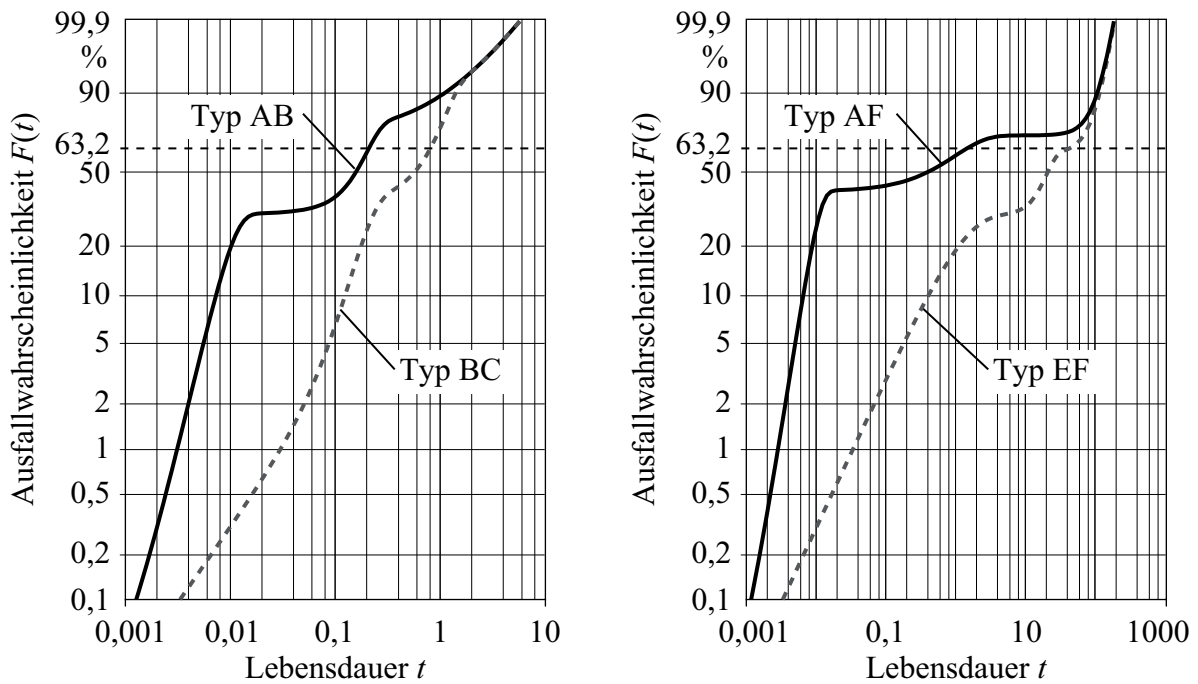


Bild 5.3: Abgeleitete Mischverteilungen mit drei Unterverteilungen

Die Separierbarkeit der Unterverteilungen nimmt von Typ AF über Typ AB und Typ EF bis zu Typ BC ab. Je stärker die Abgrenzung zwischen den Unterverteilungen der Mischverteilung ist, umso wahrscheinlicher ist eine hinreichend genaue Schätzung der Verteilungsparameter und Gewichtungsfaktoren. Die starken Abgrenzungen führen dazu, dass verschiedene Bereiche der Mischverteilung nicht von einer Unterverteilung allein abgebildet werden können. Dadurch können die Datenpunkte eindeutig einer Verteilung zugerechnet werden.

Des Weiteren werden Mischverteilungen mit vier Unterverteilungen aus den Mischverteilungstypen nach Jiang und Kececioglu generiert. Die Parameter der zugrundeliegenden Unterverteilungen sind in Tabelle 5.2 aufgeführt.

Tabelle 5.2: Parameter der abgeleiteten Mischverteilungen mit vier Unterverteilungen

	b_1	T_1	p_1	b_2	T_2	p_2	b_3	T_3	p_3	b_4	T_4	p_4
Typ ADF	1	1	0,25	3	0,01	0,25	3	4	0,25	3	100	0,25
Typ BDE	1	1	0,25	3	0,2	0,25	3	4	0,25	3	20	0,25
Typ CDE	1	1	0,25	3	1	0,25	3	4	0,25	3	20	0,25

Bild 5.4 zeigt die Mischverteilungen mit vier Unterverteilungen. Die Separierbarkeit nimmt erwartungsgemäß mit der Zunahme an Unterverteilungen ab, was sich in den gering ausgeprägten Stufen in den Verläufen zeigt.

Wie bei Mischverteilungen mit drei Unterverteilungen wurden Mischverteilungen mit unterschiedlicher Charakteristik generiert. Die Gewichtungsfaktoren p_i sind für alle Unterverteilungen gleich.

Die Verteilungen können folgendermaßen charakterisiert werden:

- Typ ADF enthält die Unterverteilungen der Typen A, D und F. Es ist ein Sattelpunkt und eine Stufe in der Mischverteilung erkennbar. Der relativ flache Verlauf der Mischverteilung zwischen dem Sattelpunkt und der Stufe ergibt sich durch den geringen Formparameter von Unterverteilung eins.
- Typ BDE ist die Kombination der Unterverteilungen der Typen B, D und E. Im Verlauf sind eine Knickstelle und eine Stufe erkennbar. Unterverteilung eins zeigt sich in dem konvexen Verlauf der Mischverteilung vor der ersten Stufe.
- Typ CDE besteht aus den Unterverteilungen der Typen C, D und E. Die Knickstelle und die Stufe im Verlauf liegen sehr eng beieinander und sind einer Geraden sehr ähnlich. Wie bei Typ BDE ist die Unterverteilung eins am konvexen Verlauf vor der Knickstelle erkennbar.

Bei den Mischverteilungen mit vier Unterverteilungen nimmt die Separierbarkeit von Typ ADF über Typ BDE zu Typ CDE ab. Eine Schätzung der Unterverteilungsparameter ist vor allem bei Typ CDE sehr schwierig.

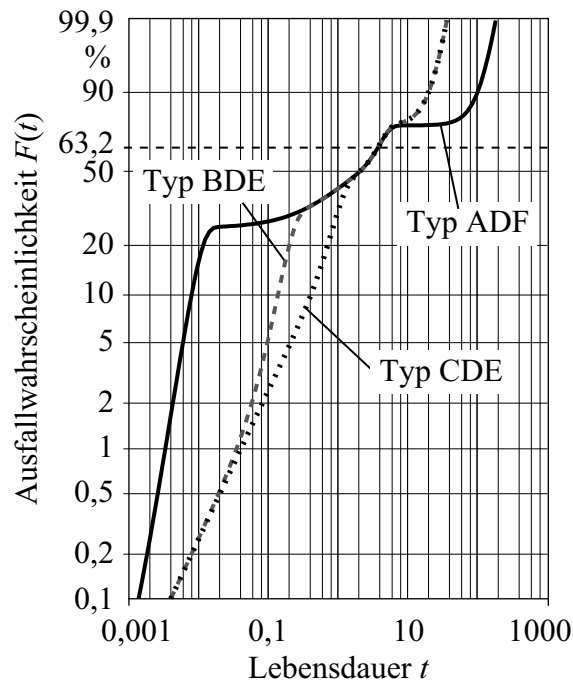


Bild 5.4: Abgeleitete Mischverteilungen mit vier Unterverteilungen

Mit der gezeigten Auswahl an Mischverteilungen aus drei und aus vier Unterverteilungen kann die Abhängigkeit der Güte der Parameterschätzung von der Separierbarkeit aufgezeigt werden. Es sind noch viele weitere kombinierte Mischverteilungen denkbar. Mit den in diesem Abschnitt gezeigten Verteilungen konnte eine Auswahl mit verschiedenen Schwierigkeitsgraden bei der Schätzung der Parameter gefunden werden.

5.1.3 Mischverteilungstypen von Jiang und Kececioglu erweitert um t_0

Die von Jiang und Kececioglu vorgestellten Mischverteilungstypen enthalten nur zweiparametrische Weibullverteilungen ohne ausfallfreie Zeit t_0 . Bei Ermüdungs- oder Verschleißausfällen tritt ein Ausfall häufig jedoch erst nach einer gewissen Laufzeit des Produkts auf. Dies ist durch physikalische Mechanismen erklärbar, bei denen eine gewisse Schädigung, die zu einem Ausfall führen kann, erst durch eine längere Laufzeit erreicht wird. Bei mechanischen Komponenten ist ein sprunghafter Anstieg, beispielsweise von Verschleiß, nicht möglich. Ein Ausfall kann also nicht schon zu Beginn der Laufzeit auftreten. Beispielsweise tritt dieses Phänomen beim Risswachstum in schwingend beanspruchten Bauteilen auf [122], weshalb bei diesen Komponenten vom Vorhandensein einer ausfallfreien Zeit ausgegangen wird.

Nach [1] besteht der in Gl. (5.4) dargestellte Zusammenhang.

$$f_{tB} = \frac{t_0}{B_{10}} \quad (5.4)$$

Darin ist f_{tB} eine vom jeweiligen Maschinenelement abhängige Größe, die B_{10} -Lebensdauer ist der Zeitpunkt, bis zu dem 10 % aller betrachteten Teile ausgefallen sind. In Tabelle 5.3 sind die f_{tB} -Faktoren für verschiedene Maschinenelemente aufgeführt. Damit lassen sich die maschinenelementabhängigen ausfallfreien Zeiten berechnen.

Tabelle 5.3: f_{tB} -Faktoren für verschiedene Maschinenelemente

Maschinenelement	Ausfallmechanismus	f_{tB}
Wellen	Bruch	0,7 ... 0,9
Wälzlager	Grübchen	0,1 ... 0,3
Zahnräder	Bruch	0,4 ... 0,8
	Grübchen	0,8 ... 0,95

Basierend auf den Werten aus Tabelle 5.3 werden die f_{tB} -Faktoren 0,1 als kleinster Wert, 0,5 als mittlerer Wert und 0,9 als Höchstwert ausgewählt. Die jeweiligen Werte für t_0 können dann über die bekannten B_{10} -Werte der Verteilungen mit Gl. (5.4) berechnet werden. Damit ergeben sich für die Mischverteilungstypen A und F aus Abschnitt 5.1.1 die in Bild 5.5 dargestellten Mischverteilungen.

Die Veränderungen am Verlauf der Mischverteilung bei Typ A und Typ F bei $f_{tB} = 0,1$ sind relativ gering, nur im Anfangsbereich der Mischverteilung ist die Verschiebung der Verteilung um t_0 nach rechts erkennbar. Bei $f_{tB} = 0,9$ ist eine klare Verschiebung um t_0 nach rechts erkennbar. Zusätzlich ändert sich der Verlauf von Typ A im Bereich des Sattelpunkts. Im Knickbereich liegt der Graph mit t_0 weiter links als der Graph

ohne t_0 . Im weiteren Verlauf im Bereich des Sattelpunktes liegt der Graph mit t_0 unterhalb des Graphen ohne t_0 . Bei Typ F bildet der Graph mit t_0 eine schärfere Knickstelle als der Graph ohne t_0 , zudem schneiden sich die Graphen bei $t \approx 100$.

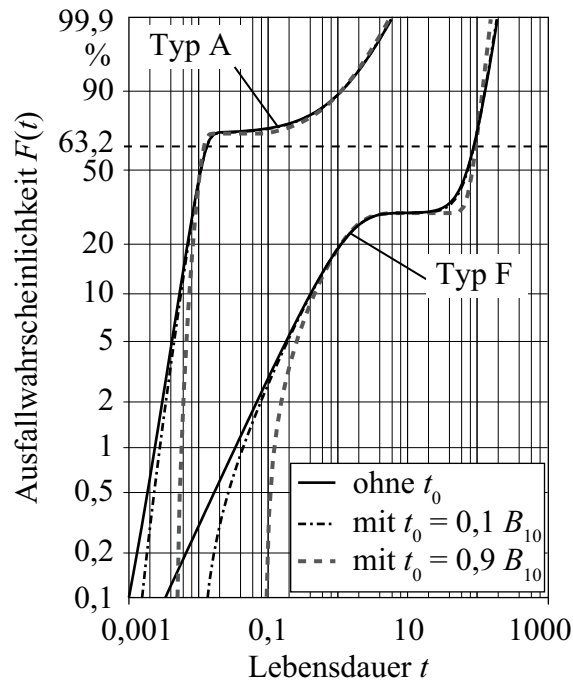


Bild 5.5: Mischverteilungen Typ A und F mit ausfallfreier Zeit t_0 ($f_{tB} = 0,1$ und $f_{tB} = 0,9$)

Es verändert sich durch die Hinzunahme der ausfallfreien Zeit der Verlauf der Mischverteilungen. Zusätzlich müssen zwei weitere Parameter der Mischverteilung geschätzt werden. Die Parameter der Datensätze mit ausfallfreier Zeit t_0 sind in Tabelle 5.4 aufgeführt. Um die Anzahl der Datensätze zu begrenzen, werden nur Typ B, Typ D und Typ F ausgewählt. Diese drei Typen repräsentieren das gesamte Spektrum der möglichen Verläufe der verschiedenen Mischverteilungen.

Tabelle 5.4: Parameter der Mischverteilungen Typ B, D, F mit ausfallfreier Zeit t_0

	b_1	T_1	t_{01}	p_1	b_2	T_2	t_{02}	p_2
Typ B _{0,1} ^{t0}	1	1	0,01	0,3	3	0,2	0,01	0,7
Typ B _{0,5} ^{t0}	1	1	0,05	0,3	3	0,2	0,05	0,7
Typ B _{0,9} ^{t0}	1	1	0,09	0,3	3	0,2	0,09	0,7
Typ D _{0,1} ^{t0}	1	1	0,01	0,3	3	4	0,19	0,7
Typ D _{0,5} ^{t0}	1	1	0,05	0,3	3	4	0,94	0,7
Typ D _{0,9} ^{t0}	1	1	0,09	0,3	3	4	1,7	0,7
Typ F _{0,1} ^{t0}	1	1	0,01	0,3	3	100	4,72	0,7
Typ F _{0,5} ^{t0}	1	1	0,05	0,3	3	100	23,62	0,7
Typ F _{0,9} ^{t0}	1	1	0,09	0,3	3	100	42,51	0,7

5.2 Boolesche Systeme

Als zweite Art der gemischten Verteilungen werden in dieser Arbeit boolesche Systeme betrachtet. Dabei muss bei den Systemen zwischen Serien- und Parallelsystemen unterschieden werden. Um die Parameter der Systeme schätzen zu können, müssen die Unterverteilungen, wie bei den Mischverteilungen, separierbar sein. Dazu muss zuerst eine Untersuchung zur Separierbarkeit durchgeführt werden.

5.2.1 Separierbarkeit

Eine Auswertung eines Datensatzes als boolesches System ist nur dann möglich, wenn sich die zugrundeliegenden Verteilungen der Komponenten einem Bereich des Ausfallverlaufs zuordnen lassen. Es muss in jedem Bereich der Systemkurve immer eine dominierende Verteilung vorhanden sein. Um einzelne Unterverteilungen trennen zu können, müssen sich die Graphen der Unterverteilungen schneiden, siehe Kapitel 2.2.3. Im Gegensatz dazu ist ein Schnittpunkt der Unterverteilungen für Mischverteilungen nicht erforderlich, beispielsweise Typ E und Typ F, siehe Kapitel 4.1. Wie bei den vorgestellten Mischverteilungen werden zuerst Systeme aus zwei Komponenten respektive mit zwei Unterverteilungen untersucht.

Die Weibullverteilung nach Gl. (2.9) aufgelöst nach der Lebensdauer t ergibt:

$$t = e^{\left(\frac{\ln(-\ln(1-F(t)))}{b}\right)} (T - t_0) + t_0 = (-\ln(1-F(t)))^{\frac{1}{b}} (T - t_0) + t_0. \quad (5.5)$$

Es wird im Folgenden nur der Fall mit $t_0 = 0$ betrachtet, damit ergibt sich aus Gl. (5.5):

$$t = T e^{\left(\frac{\ln(-\ln(1-F(t)))}{b}\right)} = T (-\ln(1-F(t)))^{\frac{1}{b}}. \quad (5.6)$$

Im Schnittpunkt der Verteilungen UV1 und UV2 gilt $t_1 = t_2 = t$ und $F_1 = F_2 = F$, damit ergibt sich für den Schnittpunkt

$$T_1 (-\ln(1-F(t)))^{\frac{1}{b_1}} = T_2 (-\ln(1-F(t)))^{\frac{1}{b_2}}. \quad (5.7)$$

Aufgelöst nach der Ausfallwahrscheinlichkeit F folgt

$$F = 1 - e^{-\left(\frac{T_1}{T_2}\right)^{\left(\frac{b_1 b_2}{b_1 - b_2}\right)}}. \quad (5.8)$$

Daraus ergibt sich die in Bild 5.6 abgebildete Schar an Grenzkurven. Dargestellt sind die Kurven der Schnittpunkte gleicher Ausfallwahrscheinlichkeiten F für den Formparameter $b_1 = 1$ in Abhängigkeit der Verhältnisse der Formparameter b_1 und b_2 und der charakteristischen Lebensdauern T_1 und T_2 . Für das Verhältnis $b_1/b_2 = 1$ ergibt sich eine Singularität durch die Parallelität der Weibullgeraden.

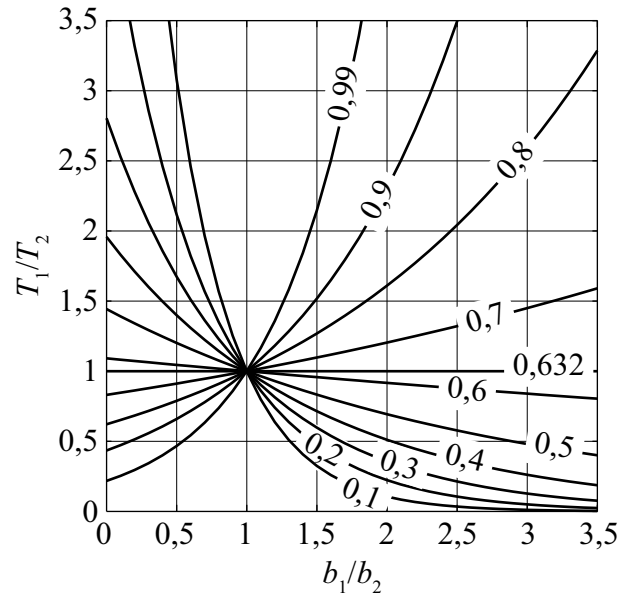


Bild 5.6: Grenzkurven für die Schnittpunkte zweier Weibullverteilungen in Abhängigkeit der Parameterverhältnisse b_1/b_2 und T_1/T_2 für $b_1 = 1$

Für $T_1/T_2 = 1$ ergeben sich die Schnittpunkte immer bei den charakteristischen Lebensdauern. Bei einem Verhältnis der Formparameter $b_1/b_2 < 1$ wandert der Schnittpunkt der Weibullgeraden mit steigendem Verhältnis T_1/T_2 zu kleineren Lebensdauern bzw. zu kleineren Ausfallwahrscheinlichkeiten. Bei $b_1/b_2 > 1$ wandert der Schnittpunkt bei kleiner werdendem Verhältnis T_1/T_2 ebenfalls zu kleineren Ausfallwahrscheinlichkeiten. Bild 5.7 zeigt die Grenzkurven für verschiedene Werte des Formparameters b_1 .

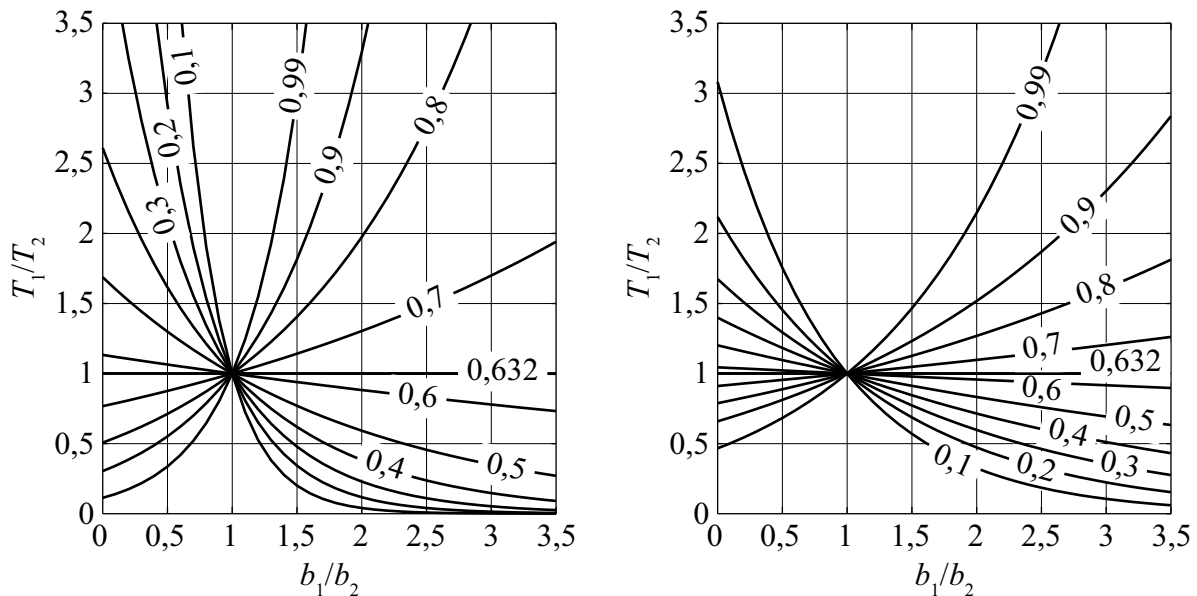


Bild 5.7: Grenzkurven für die Schnittpunkte zweier Weibullverteilungen für $b_1 = 0,7$ (links) und $b_1 = 2$ (rechts)

Bei kleineren Formparametern b_1 werden mehr Parameterkombinationen für den Formparameter b_2 und die charakteristischen Lebensdauern T_1 und T_2 möglich. Dies ist

am größeren Bereich, den die Grenzkurven einschließen, erkennbar, siehe Bild 5.7 links. Für größere Formparameter b_1 rücken die Grenzkurven näher zusammen, wodurch die Wahlmöglichkeit der Parameterkombinationen für den zweiten Formparameter und die charakteristischen Lebensdauern geringer wird, siehe Bild 5.7 rechts.

Zur Separierung der beiden Verteilungen eines Systems sind Ausfallpunkte erforderlich, die der jeweils dominierenden Ausfallverteilung zugeordnet werden können. Es muss jeweils ein Teil der Datenpunkte unterhalb und oberhalb des Schnittpunktes, aber innerhalb des durch die Daten abgedeckten Intervalls liegen. Aus der Mindestanzahl an erforderlichen Ausfallpunkten ergeben sich die oberen und unteren Grenzkurven, die den Bereich der separierbaren Parameterkombinationen einschließen. Mit dieser Überlegung folgt aus Gl. (2.5) mit der Mindestanzahl der erforderlichen Ausfallzeiten je Verteilung k mit $i = n - k$ für die obere Grenze des Schnittpunktes

$$F_{\text{grenz},o}(t) = \frac{n - k - 0,3}{n + 0,4} \quad (5.9)$$

und für die untere Grenze des Schnittpunktes

$$F_{\text{grenz},u}(t) = \frac{k - 0,3}{n + 0,4} \quad (5.10)$$

Für die Evaluation der Funktion des genetischen Algorithmus werden, wie bereits erwähnt, Datensätze mit 20, 50 und 100 Ausfallzeiten verwendet. Für $k = 5$ und $k = 10$ sind die Bereiche der Parameterkombinationen für mögliche Schnittpunkte der booleschen Systeme für $b_1 = 1$ in Bild 5.8 dargestellt.

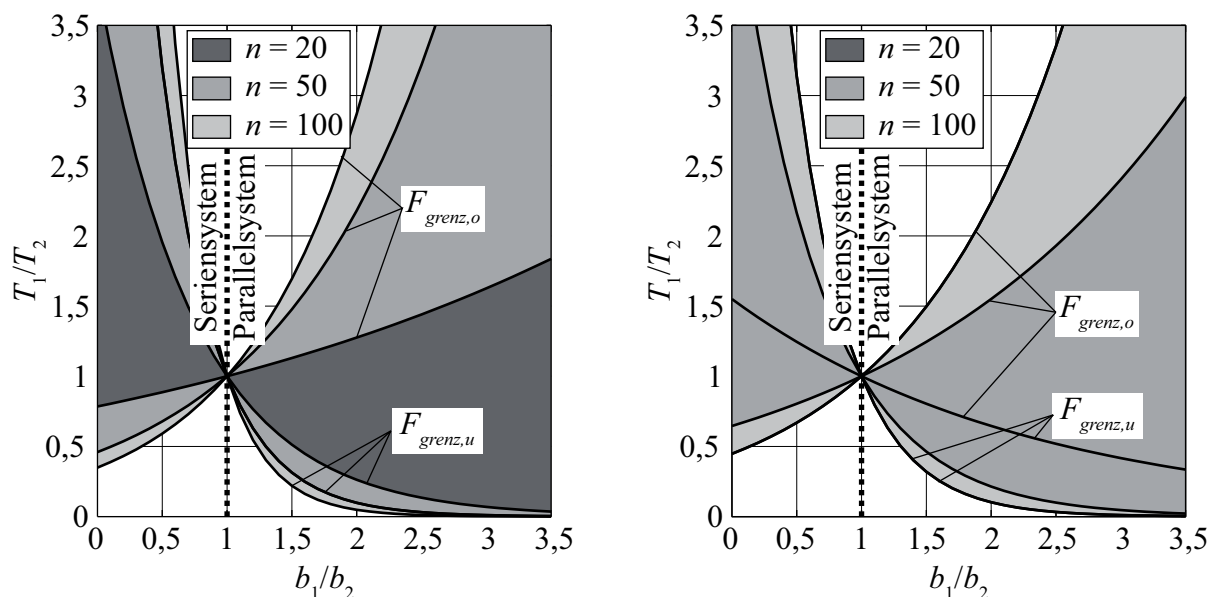


Bild 5.8: Bereiche der Parameterkombinationen für mögliche Schnittpunkte der Unterverteilungen der booleschen Systeme in Abhängigkeit der Anzahl der notwendigen Ausfallpunkte, $k = 5$ (links), $k = 10$ (rechts)

Mit steigendem Stichprobenumfang n und einer konstanten Mindestanzahl an Ausfällen je Verteilung k vergrößert sich der Bereich der separierbaren Verteilungen. Bei $n = 20$ und $k = 10$ verkleinert sich der Bereich auf eine Grenzkurve, durch die die jeweiligen Verhältnisse von b_1/b_2 und T_1/T_2 eindeutig festgelegt sind.

Die Schnittpunkte sind von der Systemstruktur unabhängig und gelten sowohl für Serien- als auch für Parallelsysteme. Als Zusatzbedingung ist $b_i < b_{i+1}$ für Seriensysteme definiert. Der kleinste Formparameter ist als b_1 definiert. Damit ist ein Schnittpunkt zwischen den Unterverteilungen nur möglich, wenn b_2 einen größeren Wert als b_1 besitzt. Die Zusatzbedingung für Parallelsysteme ist $b_i > b_{i+1}$, da der Schnittpunkt der Unterverteilungen nur auftritt, wenn der Formparameter b_2 der zweiten Verteilung kleiner ist als der Formparameter b_1 der ersten Verteilung. Dabei ist b_1 als der größte Formparameter definiert.

Um einfach separierbare boolesche Systeme zu erhalten, muss das Verhältnis b_1/b_2 für Seriensysteme möglichst klein und für Parallelsysteme möglichst groß sein. Je ähnlicher sich die Formparameter sind, umso weniger ausgeprägt ist die Knickstelle im Verlauf. Bei den von Jiang und Kececioglu vorgestellten Mischverteilungen ist ein Verhältnis b_1/b_2 von drei vorhanden [9].

Für die Bestimmung der Separierbarkeit von Systemen aus drei Komponenten sind die Grenzkurven für den Schnittpunkt der ersten und zweiten Unterverteilung mit $F_1 = F_2 = F_{1,2}$ durch Gl. (5.11) gegeben.

$$\frac{k-0,3}{n+0,4} < F_{1,2\text{grenz}}(t) < \frac{n-2k-0,3}{n+0,4} \quad (5.11)$$

Für die Grenzkurven des Schnittpunkts der zweiten und dritten Unterverteilung mit $F_2 = F_3 = F_{2,3}$ ergibt sich

$$F_{12} + \frac{k}{n+0,4} > F_{2,3\text{grenz}}(t) > \frac{n-k-0,3}{n+0,4}. \quad (5.12)$$

Mit der Definition $b_1 < b_2 < b_3$, $F_{1,2\text{grenz}} < 0,632$ ergibt sich $T_1 > T_2$ für ein Seriensystem. Für die Bestimmung der Parameterbereiche ergeben sich folgende Einschränkungen:

$$\frac{b_1}{b_2} < 1; \frac{b_2}{b_3} < 1; \frac{T_1}{T_2} > 1; \frac{T_2}{T_3} \begin{cases} > 1 & \text{für } F_{2,3} < 0,632 \\ < 1 & \text{für } F_{2,3} > 0,632 \end{cases} \quad (5.13)$$

Der Bereich möglicher Parameterkombinationen für ein Seriensystem mit drei Komponenten ist beispielhaft in Bild 5.9 (links) für $b_1 = 1$ und $b_2 = 2$ mit $n = 50$ und $k = 5$ dargestellt. Für die Darstellung wurde die untere Grenze von $F_{2,3}$ mit dem unteren Grenzwert von $F_{1,2}$ berechnet, um einen möglichst großen Bereich an

Parameterkombinationen abzubilden. Die Wahl eines größeren Wertes für $F_{1,2}$ schränkt den Bereich der wählbaren Parameterkombinationen für die dritte Unterverteilung zusätzlich ein.

Für Parallelsysteme mit drei Unterverteilungen gelten ebenfalls die Grenzen aus Gl. (5.11) und Gl. (5.12). Für die Parameter der Unterverteilungen gelten die Beschränkungen: $b_1 > b_2 > b_3$, $F_{1,2,grenz} < 0,632$ und damit $T_1 < T_2$. Die Beschränkungen der Parameterbereiche ergeben sich mit

$$\frac{b_1}{b_2} > 1; \frac{b_2}{b_3} > 1; \frac{T_1}{T_2} < 1; \frac{T_2}{T_3} \begin{cases} < 1 & \text{für } F_{2,3} < 0,632 \\ > 1 & \text{für } F_{2,3} > 0,632 \end{cases} \quad (5.14)$$

Der Bereich der möglichen Parameterkombinationen für ein beispielhaftes Parallelsystem mit drei Komponenten ist in Bild 5.9 (rechts) für $b_1 = 1$ und $b_2 = 0,7$ mit $n = 50$ und $k = 5$ dargestellt. Für die Darstellung wurde wie beim Seriensystem die untere Grenze von $F_{2,3}$ mit dem unteren Grenzwert von $F_{1,2}$ berechnet, um auch hier einen möglichst großen Bereich der Parameterkombinationen abzubilden.

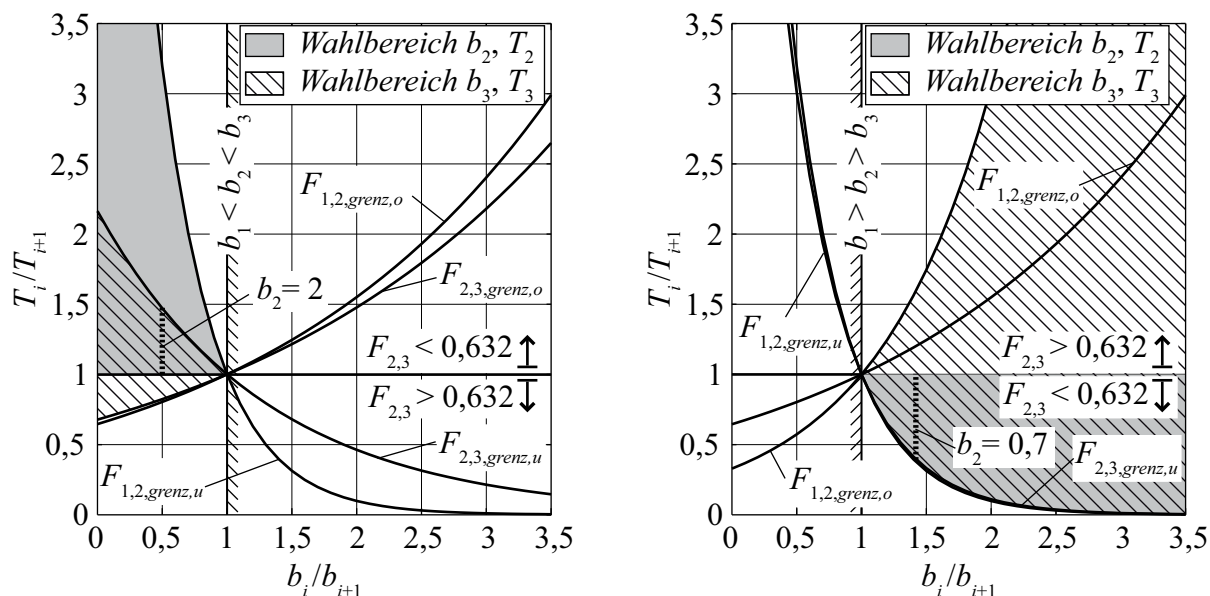


Bild 5.9: Bereiche der Parameterkombinationen für mögliche Schnittpunkte der Unterverteilungen der booleschen Serien- (links) und Parallelsysteme (rechts) beispielhaft für $b_1 = 1$ und $b_2 = 2$ bzw. $b_2 = 0,7$ mit $n = 50$ und $k = 5$

Im Übergangsbereich zwischen den Unterverteilungen von Systemen tritt keine scharfe Knickstelle auf, was bedeutet, dass der Schnittpunkt nicht eindeutig identifizierbar ist. Vor allem bei Parallelsystemen ist dieser sanfte Übergang der Unterverteilungen innerhalb der Systemkurve zu beachten, siehe Bild 2.5. Die Separierbarkeit einzelner Unterverteilungen voneinander wird somit erschwert. Zusätzlich wird der Bereich der möglichen Parameterkombinationen weiter beschränkt, da pro Verteilung mehr Ausfallzeiten und damit ein größeres k erforderlich sind.

5.2.2 Datensätze der booleschen Seriensysteme

Für die Betrachtung von Seriensystemen ist der Bereich in Bild 5.8 mit $b_1/b_2 < 1$ entscheidend, da $b_1 > b_2$ definiert ist. Ausgehend von den Überlegungen zur Separierbarkeit werden die Parameter b_1 und b_2 zuerst festgelegt. Für Seriensysteme wird für T_1 ein Wert von 100 definiert. Aus diesen Angaben kann über Gl. (5.15) die Zeit T_2 in Abhängigkeit von b_1 , b_2 , T_1 und der Wahrscheinlichkeit $F_{1,2}$ am Schnittpunkt berechnet werden.

$$T_2 = T_1(-\ln(1 - F_{1,2}))^{\left(\frac{b_2 - b_1}{b_1 b_2}\right)} \tag{5.15}$$

Um die Parameterschätzungen mit verschiedenen Stichprobenumfängen realisieren zu können, dienen die obere und untere Grenzkurve von $n = 20$ und $k = 5$ als Basis. Damit sind Parameterkombinationen auch bei Datensätzen mit Stichprobenumfängen von $n = 50$ und $n = 100$ separierbar. Zusätzlich ist als dritte Variante die Ausfallwahrscheinlichkeit bei einem Schnittpunkt mit $F(t) = 50\%$ definiert. Diese Variante liegt genau zwischen den Grenzkurven, wodurch eine bessere Separierbarkeit erreicht werden kann. Die Grenzkurven und die gewählten Seriensysteme zeigt Bild 5.10.

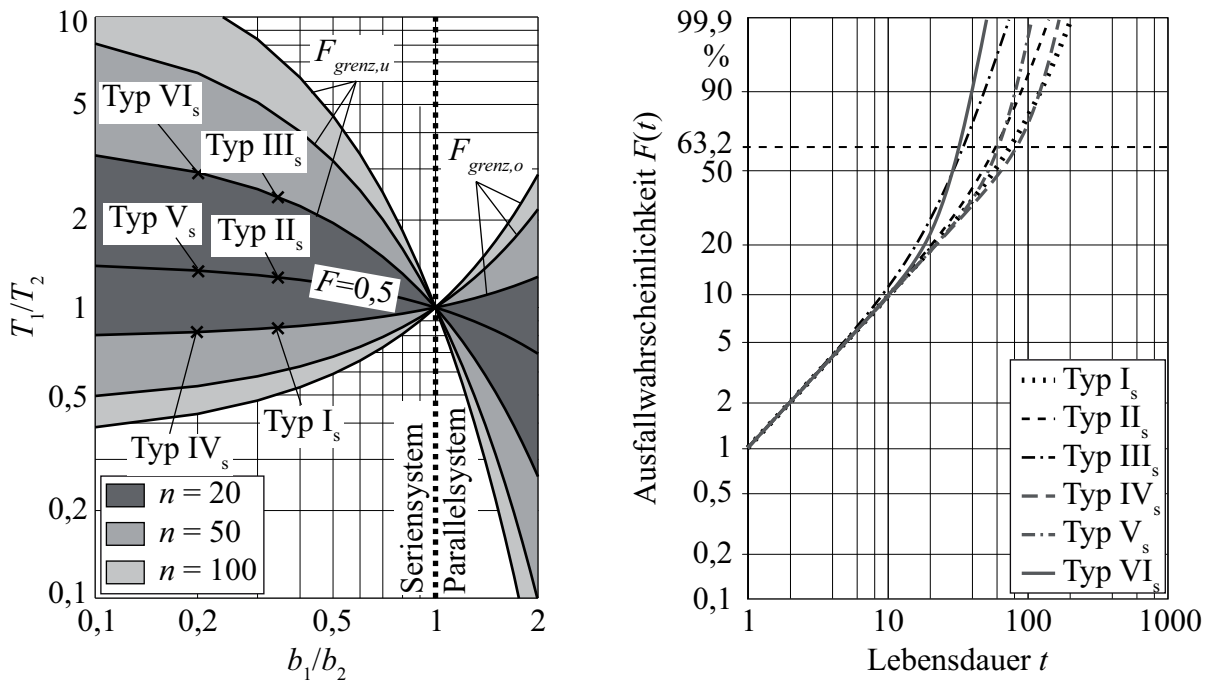


Bild 5.10: Gewählte boolesche Seriensysteme aus zwei Komponenten Typ I_s – Typ VI_s mit $b_1 = 1$ im Grenzkurvenbild (links) und als Systemkurven (rechts)

In Anlehnung an die Mischverteilungen werden die verschiedenen Systeme als Typ I_s bis Typ VI_s bezeichnet. Der Index s kennzeichnet die Seriensysteme. Durch die Wahl der Systeme Typ I_s und Typ IV_s auf der oberen Grenzkurve und Typ III_s und Typ VI_s auf der unteren Grenzkurve kann die Abhängigkeit der Güte der Schätzung von der

Lage des Schnittpunktes untersucht werden. Zusätzlich wird der Formparameter b_2 variiert. Die Systemkurven für die gewählten Typen I_s bis VI_s sind in Bild 5.10 (rechts) dargestellt. Erkennbar sind die Änderungen der Steigung der Systemkurve bei den Typen III_s und VI_s bei einer Ausfallwahrscheinlichkeit von etwa 20 %. Bei den Typen I_s und IV_s findet die Änderung der Steigung bei ungefähr 60 % Ausfallwahrscheinlichkeit statt. Die Typen II_s und V_s liegen dazwischen. In Tabelle 5.5 sind die Parameter für die booleschen Seriensysteme aufgeführt. Die Formparameter b_2 sind so gewählt, dass eine eindeutige Separierbarkeit möglich ist.

Tabelle 5.5: Parameter der booleschen Seriensysteme Typ I_s – Typ VI_s

	b_1	T_1	b_2	T_2
Typ I_s	1	100	3	120
Typ II_s	1	100	3	80
Typ III_s	1	100	3	40
Typ IV_s	1	100	5	120
Typ V_s	1	100	5	75
Typ VI_s	1	100	5	35

Zur Untersuchung der ausfallfreien Zeit t_0 werden, wie bei den Mischverteilungen nach Jiang und Kececioglu, die Parameterkombinationen der Seriensysteme genutzt. Zur Einschränkung der Vielfalt an Kombinationen werden die am besten separierbaren Systeme Typ II_s und Typ V_s ausgewählt. Die Seriensysteme werden dazu mit den nach Gl. (5.4) über den f_{iB} -Wert berechneten t_0 -Werten erweitert. Die f_{iB} -Werte werden mit 0,1 und 0,9 angenommen. Tabelle 5.6 zeigt die Parameter der booleschen Seriensysteme mit ausfallfreier Zeit t_0 .

Tabelle 5.6: Parameter der booleschen Seriensysteme mit ausfallfreier Zeit t_0

	b_1	T_1	t_{01}	b_2	T_2	t_{02}
Typ $II_{s,0,1}^{10}$	1	100	1,1	3	80	3,8
Typ $II_{s,0,9}^{10}$	1	100	5,3	3	80	18,9
Typ $V_{s,0,1}^{10}$	1	100	1,1	5	75	4,8
Typ $V_{s,0,9}^{10}$	1	100	5,3	5	75	23,9

5.2.3 Datensätze der booleschen Parallelsysteme

Das Vorgehen bei der Auswahl der booleschen Parallelsysteme ist analog zur Auswahl der Seriensysteme. Um Parameterschätzungen mit kleinen Stichproben durchführen zu können, dienen die Grenzkurven mit $n = 20$ und $k = 5$ und die Grenzkurve für den Schnittpunkt bei 50 % Ausfallwahrscheinlichkeit als Basis für die Auswahl der Parameter der Unterverteilungen. Die Grenzkurven für die Parallelsysteme und die gewählten Parallelsystemtypen sind in Bild 5.11 dargestellt.

Zur besseren Unterscheidbarkeit sind die Parallelsystemtypen mit dem Index p mit Typ I_p bis Typ VI_p bezeichnet. Ebenso wie bei den Seriensystemen zeigt sich eine Änderung der Steigung der Systemkurven. Erkennbar ist dies bei den Typen III_p und VI_p bei einer Ausfallwahrscheinlichkeit von etwa 80 %. Bei den Typen I_p und IV_p findet die Änderung der Steigung bei ungefähr 30 % Ausfallwahrscheinlichkeit statt. Die Typen II_p und V_p liegen zwischen den anderen Systemkurven.

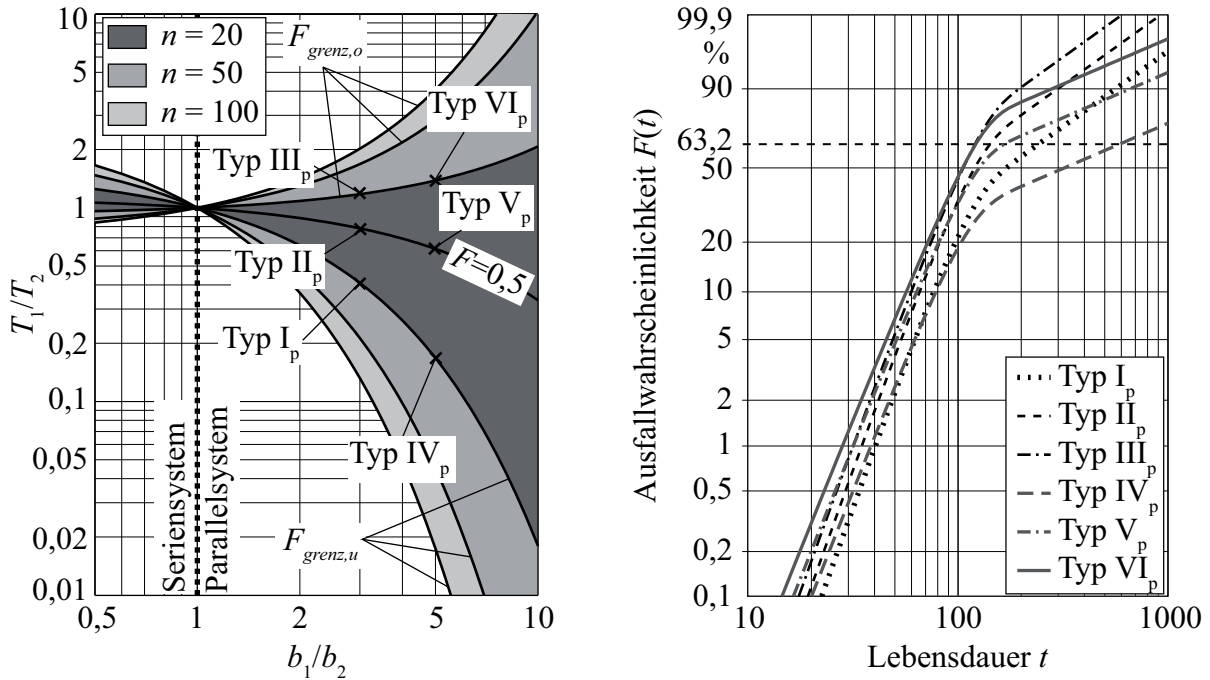


Bild 5.11: Gewählte boolesche Parallelsysteme aus zwei Komponenten Typ I_p – Typ VI_p mit $b_1 = 3$ im Grenzkurvenbild (links) und als Systemkurven (rechts)

Der Formparameter der ersten Unterverteilung ist wie bei den Seriensystemen konstant mit $b = 3$ gewählt. Für die zweite Unterverteilung variiert der Wert des Formparameters zwischen 0,6 und 1. Die Werte für die charakteristische Lebensdauer können nach Gl. (5.15) berechnet werden.

Die Parameter der gewählten booleschen Parallelsysteme sind in Tabelle 5.7 aufgeführt.

Tabelle 5.7: Parameter der booleschen Parallelsysteme Typ I_p – Typ VI_p

	b_1	T_1	b_2	T_2
Typ I_p	3	100	1	245
Typ II_p	3	100	1	130
Typ III_p	3	100	1	85
Typ IV_p	3	100	0,6	600
Typ V_p	3	100	0,6	165
Typ VI_p	3	100	0,6	70

In Tabelle 5.8 sind die Parameter für die booleschen Parallelsysteme mit ausfallfreier Zeit t_0 dargestellt. Die Grundlage hierzu sind die Parallelsystemtypen II_p und V_p mit einem f_{iB} -Wert von 0,1 und 0,5.

Tabelle 5.8: Parameter der booleschen Parallelsysteme mit ausfallfreier Zeit t_0

	b_1	T_1	t_{01}	b_2	T_2	t_{02}
Typ $\text{II}_{p,0,1}^{t_0}$	3	100	4,7	1	130	1,4
Typ $\text{II}_{p,0,5}^{t_0}$	3	100	23,6	1	130	6,8
Typ $\text{V}_{p,0,1}^{t_0}$	3	100	4,7	0,6	165	0,4
Typ $\text{V}_{p,0,5}^{t_0}$	3	100	23,6	0,6	165	1,9

5.3 Zensierte Datensätze

Bei der Auswertung von Felddaten ist eine unvollständige Stichprobe die Regel [2]. Es werden also Stichproben, die nicht ausgefallene Teile bzw. Teile mit unbekanntem Lebensdauermerkmalen enthalten, auftreten. Diese unbekanntem Daten verursachen Lücken im Verlauf der Ausfalldaten. Die Zensierung kann deshalb eine Verschlechterung der Separierbarkeit der vorhandenen Unterverteilungen bewirken. Dies ist der Fall, wenn signifikante Punkte im Ausfallverlauf, wie beispielsweise Knickstellen, Stufen oder Sattelpunkte, nicht mehr eindeutig erkennbar sind. Zudem kann die Lage von Knickstellen durch die Zensierung der Daten sowohl in Ordinatens- als auch in Abszissenrichtung verändert werden. Dadurch ist eine genaue Abgrenzung der Unterverteilungen nicht mehr möglich und es resultieren daraus Parameterschätzungen mit schlechterer Güte.

Bei der Parameterschätzung mit genetischen Algorithmen werden die gegebenen Datenpunkte möglichst gut durch eine Verteilungskurve angenähert. Als Eingangswerte dienen die Ausfalldaten mit der Ausfallzeit und der zugeordneten Ausfallwahrscheinlichkeit. Für zensierte Daten müssen die Ausfallwahrscheinlichkeiten beispielsweise über das Johnson Verfahren oder über den Kaplan-Meier-Schätzer (siehe Kapitel 2.1) bestimmt werden, um die unvollständigen Daten zu berücksichtigen.

Für die Parameterschätzung mit genetischen Algorithmen ist von besonderem Interesse, ob die zensierten Daten die Ausfallverteilung der vollständigen Stichprobe repräsentativ abbilden. Ist dies der Fall, sind Parameterschätzungen von ähnlicher Güte wie bei vollständigen Stichproben zu erwarten. Das prinzipielle Vorgehen zur Parameterschätzung entspricht, bis auf die Bestimmung der Ausfallwahrscheinlichkeiten, dem Ablauf bei vollständigen Stichproben. Daher ist eine getrennte Betrachtung von zensierten Datensätzen nicht erforderlich.

6 Parameterschätzung von gemischten Verteilungen

In den folgenden Abschnitten werden Parameterschätzungen von gemischten Verteilungen mit genetischen Algorithmen durchgeführt. In Kapitel 5 wurden repräsentative gemischte Verteilungen vorgestellt, anhand derer die Anwendbarkeit des genetischen Algorithmus auf verschiedene Problemstellungen untersucht wird. Darüber hinaus wird für eine dreiparametrische Weibullverteilung ein Vergleich mit verschiedenen analytischen Parameterschätzverfahren gezeigt.

Die Untersuchungen der verschiedenen Parameter des genetischen Algorithmus in den vorigen Kapiteln zeigen die Notwendigkeit einer geeigneten Parametrierung der Initialisierungsgrößen. Die Definition der Fitnessfunktion mit einer Berechnung in Ordinateurichtung und einer Gewichtung in Weibullkoordinaten, die Rekombination über N-Punkt-Crossover, die Mutationswahrscheinlichkeit nach Jansen und die Selektionsvariante Var_1 mit 20 % Elternindividuen werden für die folgenden Parameterschätzungen aus Kapitel 4.4.1 bis Kapitel 4.4.3 übernommen. Die Populationsgröße, die Größe des Lösungsraumes und die Genauigkeit der Parameter werden problemspezifisch angepasst. Die jeweiligen verwendeten Parameter sind im Anhang A1 angegeben.

Alle verwendeten Datensätze wurden anhand der Inversionsmethode über die Monte-Carlo Methode erzeugt. Die so entstandenen Datensätze sind im Anhang A2 aufgeführt.

6.1 Methode zum Vergleich der geschätzten Verteilungen

Um die Ergebnisse der Parameterschätzungen miteinander verglichen zu können, muss eine Größe verwendet werden, die eine Möglichkeit bietet, die Abweichungen zu den gegebenen Datensätzen zu erkennen. Ein Vergleich der einzelnen Parameter untereinander ist nicht zielführend, da jeder Parameter einer Unterverteilung eine Veränderung der Gesamtcharakteristik der Mischverteilung oder des Systems bewirken kann. Eine gute Übereinstimmung eines Parameters mit den Vorgabewerten der Monte-Carlo Methode ist damit nicht zwangsläufig ein Indiz für eine gute Schätzung der gesamten Verteilung. Deshalb sind zur Feststellung der Güte nichtparametrische Tests geeignet [123], [124], [125].

Der Kolmogorov-Smirnov-Test ist ein nichtparametrischer statistischer Test, mit dem die Anpassung einer theoretischen Verteilungsfunktion an eine empirische Verteilungsfunktion überprüft werden kann [126], [127]. Er ist insbesondere für kleine Stichproben geeignet und verteilungsunabhängig. Die Anwendung des Kolmogorov-

Smirnov-Tests auf Mischverteilungen wurde in [128], [129], [130], [131] und [132] untersucht.

Beim Kolmogorov-Smirnov-Anpassungstest wird die Nullhypothese H_0 nach Gl. (6.1) gegen die Alternativhypothese H_A nach Gl. (6.2) überprüft.

$$H_0 : F_n(x) = F_0(x) \quad (6.1)$$

$$H_A : F_n(x) \neq F_0(x) \quad (6.2)$$

Darin ist F_n die empirische, tatsächlich zugrunde liegende Verteilungsfunktion und F_0 die theoretische Verteilungsfunktion. Der maximale Ordinatenabstand D_{max} zwischen der Stichprobe $F_B = F_n$ und der theoretischen Verteilungsfunktion $F_E = F_0$, bezogen auf den Stichprobenumfang n , wird mit Gl. (6.3) zum Nachweis der Nullhypothese bestimmt.

$$D_{max} = \frac{\max |F_B - F_E|}{n} \quad (6.3)$$

Die Nullhypothese wird abgelehnt, wenn D_{max} einen kritischen Wert D_{krit} überschreitet. Der kritische Wert wird über das Signifikanzniveau oder Irrtumswahrscheinlichkeit α bestimmt und kann aus Tabellenwerken entnommen werden [29]. Für verschiedene Stichprobenumfänge sind die Werte für D_{krit} , bezogen auf den Stichprobenumfang n , für die Signifikanzniveaus $\alpha = 5\%$ und $\alpha = 10\%$ in Tabelle 6.1 aufgeführt.

Tabelle 6.1: Kritische Werte D_{krit} bezogen auf den Stichprobenumfang n für den Kolmogorov-Smirnov-Anpassungstest

Stichprobenumfang n	kritische Werte D_{krit}	
	$D_{0,1}$	$D_{0,05}$
20	0,01325	0,0147
50	0,0034	0,00376
100	0,00121	0,00134

In Anlehnung an den Kolmogorov-Smirnov-Anpassungstest werden zum Vergleich der verschiedenen Parameterschätzungen die Ordinatenabstände zwischen den geschätzten Verteilungen und den Medianrängen der Datenpunkte verwendet. Dazu werden sowohl die Maximalwerte der Abweichungen D_{max} als auch die Medianwerte D_{median} der Einzelabweichungen $D_x = |F_B - F_E|$, bezogen auf den Stichprobenumfang, bestimmt. Damit ist es möglich, einen Eindruck von der maximalen Abweichung der geschätzten Verteilung und einen Eindruck von der mittleren Approximation der Datenpunkte zu erhalten. Durch die Bestimmung der Abweichungen, bezogen auf den Stichprobenumfang, ist es zusätzlich möglich, die Werte von D_{max} und D_{median} für unterschiedliche Stichprobengrößen miteinander zu vergleichen. In beiden Fällen ist die

Annäherung an die Daten besser, je kleiner D_{max} und D_{median} sind. Diese Herangehensweise eignet sich auch zur Auswertung von Parameterschätzungen mit analytischen Methoden, siehe Kapitel 2.3, mit den zu schätzenden Datenpunkten. Damit können die Ergebnisse der verschiedenen Methoden verglichen werden.

6.2 Parameterschätzung von Einfachverteilungen

In einem ersten Schritt werden die Ergebnisse des genetischen Algorithmus mit den Ergebnissen von drei verschiedenen analytischen Verfahren verglichen, um die prinzipielle Eignung zur Parameterschätzung von Verteilungen zu untersuchen. Zudem wird der Einfluss der Stichprobengröße auf die Parameterschätzung betrachtet. Als analytische Methoden kommen die Regression, die Maximum-Likelihood Methode und die Momentenmethode zum Einsatz. Die gewählte Verteilung ist eine dreiparametrische Weibullverteilung mit einem Formparameter $b = 1$ und einer charakteristischen Lebensdauer $T = 100$. Die ausfallfreie Zeit t_0 wurde über die f_{tB} -Werte 0,5 und 0,9 variiert. Die Datensätze mit drei verschiedenen Stichprobengrößen n von 20, 50 und 100 wurden mit Hilfe der Monte-Carlo Methode erzeugt. Die verwendeten Datensätze sind im Anhang in Tabelle A2.1 bis Tabelle A2.9 und die Parameter des Algorithmus in Tabelle A1.7 und Tabelle A1.8 zu finden.

Die Ergebnisse der Parameterschätzungen mit analytischen Methoden und mit genetischem Algorithmus sind in Bild 6.1 dargestellt.

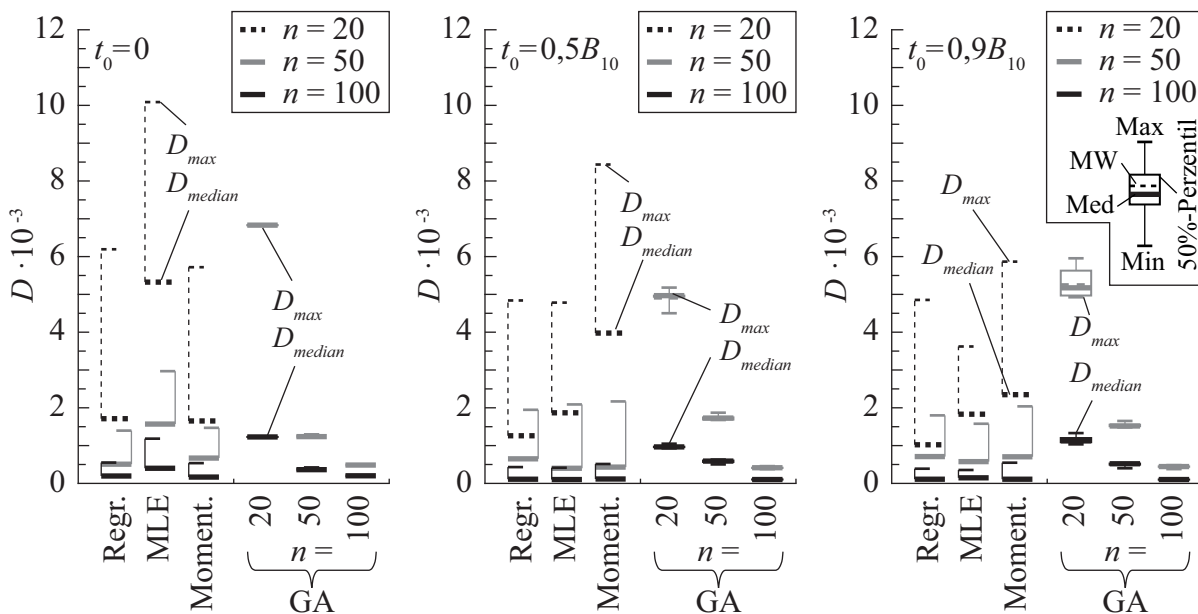


Bild 6.1: Ergebnisse der Parameterschätzungen der dreiparametrischen Weibullverteilung mit $t_0 = 0$ (links), $t_0 = 0,5B_{10}$ (Mitte) und $t_0 = 0,9B_{10}$ (rechts)

Dargestellt sind in Bild 6.1 die Ergebnisse der analytischen Methoden, jeweils mit verschiedenen Stichprobenumfängen n für Weibullverteilungen mit unterschiedlichen

Werten für t_0 . Aufgetragen ist für jedes Ergebnis die Spannweite zwischen dem maximalen Ordinatenabstand D_{max} und dem Median des Ordinatenabstands D_{median} . Bei allen Parameterschätzungen liegen die Abweichungen in ähnlichen Größenordnungen. Mit steigender Stichprobengröße nehmen für alle Methoden die Werte für D_{max} und D_{median} ab.

Im Bild 6.1 sind die erzielten Werte des genetischen Algorithmus für D_{max} und D_{median} als Box-Plot für jeweils 20 Replikationen je Verteilung über den Stichprobenumfang aufgetragen. Die Werte liegen in einem ähnlichen Bereich wie die der analytischen Methoden, teilweise etwas besser. Die Schätzungen der Parameter derselben Verteilungen sind bei allen 20 Replikationen sehr ähnlich, was sich in den sehr schmalen Box-Plots zeigt. Die Güte der Schätzung steigt auch beim genetischen Algorithmus mit der Zunahme der Stichprobengröße an.

Es konnte damit für die beispielhafte Verteilung gezeigt werden, dass mit einem genetischen Algorithmus Parameterschätzungen von Weibullverteilungen möglich sind. Die Ergebnisse sind mit denen der bekannten analytischen Verfahren vergleichbar.

6.3 Parameterschätzung von Mischverteilungen

In diesem Abschnitt werden die Parameterschätzungen der in Kapitel 5.1 vorgestellten Mischverteilungen durchgeführt und die Ergebnisse verglichen. Als Vergleichsgröße dienen der maximale Ordinatenabstand D_{max} und der Median der Ordinatenabstände D_{median} . Zuerst werden die aus Kapitel 4.1 bekannten Mischverteilungen nach Jiang und Kececioglu untersucht. Anschließend werden die Parameterschätzungen zu den daraus abgeleiteten Mischverteilungen mit mehreren Unterverteilungen sowie ausfallfreier Zeit durchgeführt.

Alle Datensätze wurden mittels Monte-Carlo Methode erzeugt und finden sich im Anhang A2. Die jeweils verwendeten Parameter des genetischen Algorithmus finden sich ebenfalls im Anhang A1.

6.3.1 Parameterschätzung der Mischverteilungen nach Jiang und Kececioglu

Die Parameterschätzung der Mischverteilungen nach Jiang und Kececioglu wurde mit den in Tabelle A1.9 aufgeführten Parametern für den genetischen Algorithmus durchgeführt. Die Datensätze finden sich im Anhang Tabelle A2.10 bis Tabelle A2.21. Für jeden Datensatz wurden 20 Replikationen durchgeführt. Zum Vergleich der Ergebnisse dienen die Schätzungen der Mischverteilungen mit der Software Weibull++[®] von Reliasoft[®]. Der Mischverteilungsschätzer von Weibull++[®] nutzt den MLE/EM-Algorithmus, siehe Kapitel 2.3.3, der über einen Levenberg-Marquardt-Algorithmus gelöst wird [133]. Der Levenberg-Marquardt-Algorithmus ist ein numerisches

Optimierungsverfahren, das den gradientenbasierten, deterministischen Methoden zugeordnet werden kann [134].

Die Ergebnisse der Parameterschätzungen des genetischen Algorithmus und des MLE/EM-Algorithmus in Weibull++[®] sind in Bild 6.2 dargestellt.

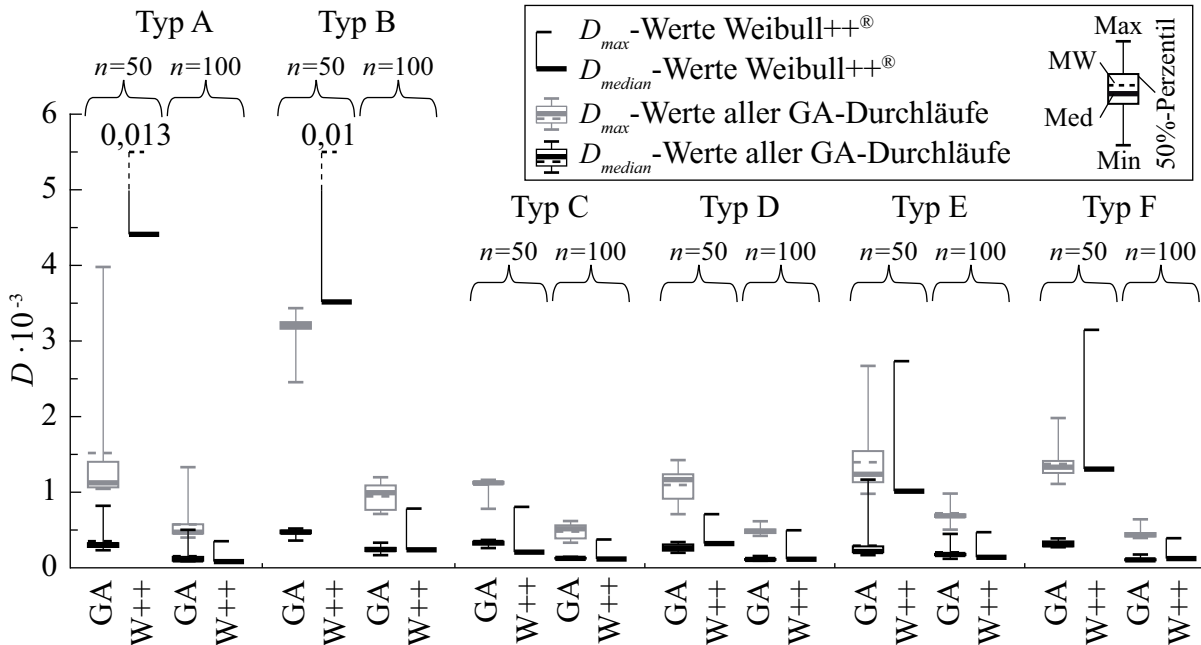


Bild 6.2: Ergebnisse der Parameterschätzungen der Mischverteilungen nach Jiang mit Weibull++[®] und genetischem Algorithmus

Aufgetragen sind der maximale Ordinatenabstand D_{max} und der Median aller Ordinatenabstände D_{median} für die Schätzung mittels MLE/EM-Algorithmus. Für die Schätzungen mittels genetischem Algorithmus sind die Box-Plots für die Werte von D_{max} und D_{median} für jeweils 20 durchgeführte Replikationen dargestellt.

Für Typ A zeigen beide Methoden die schlechtesten Ergebnisse. Vor allem die MLE/EM-Methode aus Weibull++ zeigt sehr große Werte für D_{max} . Hier ist die Schwierigkeit in dem scharfen Knick vor dem Sattelpunkt zu sehen, der durch den streuenden Datensatz nicht eindeutig definiert ist. Daraus resultieren große Abweichungen der Ordinaten bei nur kleinen Abweichungen der geschätzten Parameter. Dasselbe Problem tritt bei Typ B bei geringen Stichprobenumfängen auf. Auch hier verursacht eine kleine Abweichung der geschätzten Parameter eine große Abweichung der Ordinatenwerte im Bereich der Knickstelle.

Die Mischverteilungstypen C und D erzielen bei der Parameterschätzung durch das Fehlen von scharfen Übergängen die besten Ergebnisse, was sich in geringen Abweichungen der Ordinaten bei beiden Stichprobenumfängen zeigt. Zudem liegen die charakteristischen Lebensdauern der Verteilungen innerhalb ähnlicher Größenordnungen, was sich günstig auf die mittels Rekombination erzeugten Nachkommen auswirkt. Die

Werte von T sind beispielsweise bei Mischverteilungstyp C identisch und damit austauschbar.

Bei den Mischverteilungstypen E und F ist wieder ein deutlicher Anstieg der Abweichungen bei kleinen Stichprobenumfängen zu erkennen. Bei Typ E ist dies auf die eng beieinander liegenden Knickstellen zurückzuführen, die im mit der Monte-Carlo Methode erzeugten Datensatz nicht eindeutig trennbar sind. Die Stufe im Verlauf verursacht bei Typ F wie bei Typ A eine kritische Stelle bei der Parameterschätzung, wodurch die höheren Abweichungen bei geringen Stichprobenumfängen erklärt werden können.

Die Verteilungen der Parameterschätzungen sind in Bild 6.3 für den Mischverteilungstyp A mit Stichprobenumfang $n = 50$ und Typ F mit $n = 100$ beispielhaft dargestellt.

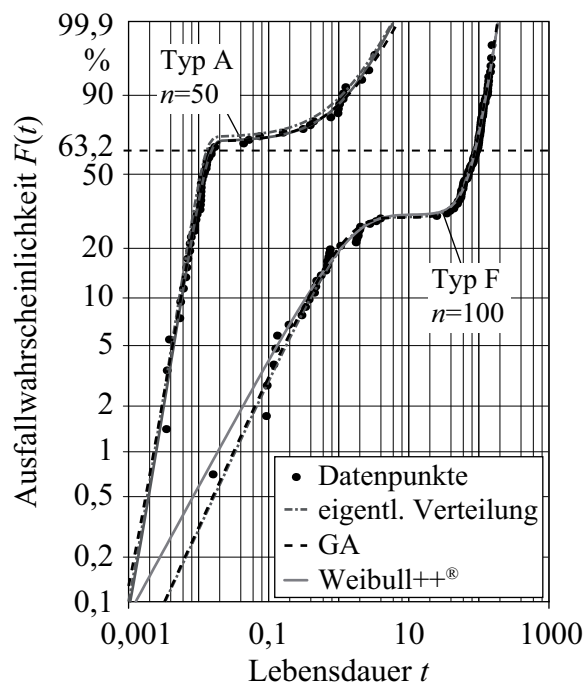


Bild 6.3: Geschätzte Verteilungsfunktionen für Typ A ($n = 50$) und Typ F mit ($n = 100$) mit Weibull++[®] und genetischem Algorithmus

Die dargestellten Verläufe der mittels genetischem Algorithmus geschätzten Parameter gehen aus den Schätzungen mit den besten Fitnesswerten aus den 20 Replikationen hervor. Es sind die Typen mit sehr guten (Typ F) und sehr schlechten (Typ A) Werten für die Ordinatenabstände gezeigt. Die mittels Monte-Carlo Methode simulierten Daten geben den Verlauf der Verteilungen sehr gut wieder.

Zu erkennen sind die wenigen Datenpunkte im Bereich der Sattelpunkte der Verteilungen, wodurch die Annäherung an diesen Bereich erschwert wird. Die geschätzten Verteilungen sind nahezu deckungsgleich mit den Datenpunkten, selbst bei Mischverteilungstyp A mit großen Werten für D_{max} und D_{median} .

Bei den Parameterschätzungen der Mischverteilungen nach Jiang und Kececioglu konnten mit dem genetischen Algorithmus durchweg gute Ergebnisse erzielt werden. Die Werte liegen im Bereich der mittels der kommerziellen Software Weibull++[®] erreichten Werte. Dies gilt sowohl für die erreichbaren Werte von D_{max} als auch D_{median} .

6.3.2 Parameterschätzung von Mischverteilungen mit mehreren Unterverteilungen

Im Folgenden werden die in Kapitel 5.1.2 vorgestellten Datensätze mit mehreren Unterverteilungen betrachtet. Bei diesen Datensätzen wurden die Mischverteilungen von Jiang und Kececioglu um weitere Unterverteilungen erweitert. Zuerst werden Datensätze mit drei Unterverteilungen betrachtet, siehe Tabelle A2.22 bis Tabelle A2.25. Die zur Parameterschätzung verwendeten Parameter des Algorithmus finden sich in Tabelle A1.10. Der Stichprobenumfang liegt für alle Verteilungen bei $n = 100$.

Die Ergebnisse der Parameterschätzungen der erweiterten Mischverteilungen mit drei Unterverteilungen mit der Software Weibull++[®] und dem genetischen Algorithmus sind in Bild 6.4 dargestellt.

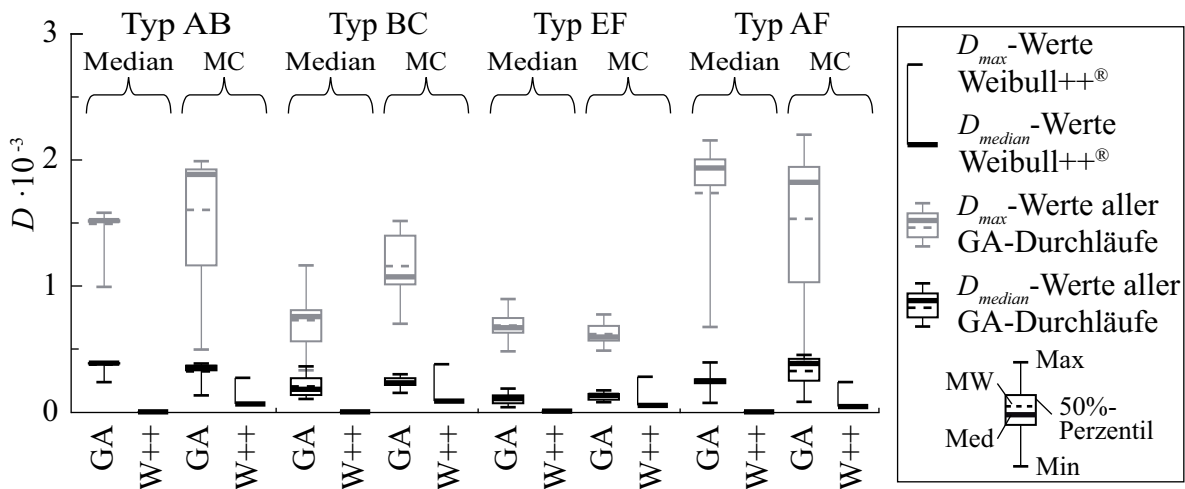


Bild 6.4: Ergebnisse der Parameterschätzungen der Mischverteilungen mit drei Unterverteilungen mit Weibull++[®] und genetischem Algorithmus

Aufgetragen sind die Werte für D_{max} und D_{median} , die mit Weibull++[®] berechnet wurden, sowie die Box-Plots für die entsprechenden Werte, die aus 20 Replikationen des genetischen Algorithmus resultieren. Für alle Verteilungstypen mit drei Unterverteilungen wurden sowohl Datensätze mit Medianwerten als auch mittels Monte-Carlo Methode erzeugte Datensätze untersucht. Dadurch können die Auswirkungen der stochastischen Abweichungen von der idealen Verteilung auf die Parameterschätzung untersucht werden.

Deutlich zu erkennen ist eine Zunahme der Werte von D_{max} und D_{median} gegenüber den Schätzungen der Verteilungen mit zwei Unterverteilungen beim genetischen Algorithmus. Dies ist vor allem auf die drei zusätzlich zu schätzenden Parameter für die dritte Unterverteilung zurückzuführen. Hierdurch wird die Wahrscheinlichkeit eines Individuums im Bereich der gesuchten Lösung bei gleicher Lösungsraumgröße, Genauigkeit und Populationsgröße verringert, siehe Gl. (4.27).

Sehr deutlich zeigen sich diese großen maximalen Abweichungen bei den Typen AB und AF. Bei den Parameterschätzungen mit den Datensätzen mit Medianwerten liegt der Medianwert bei einem sehr kleinen 50 %-Perzentil sehr hoch, was auf ein sehr starkes lokales Optimum bei einer relativ großen Abweichung schließen lässt. Bei beiden Typen AB und AF ist die erreichbare Güte kaum von der Art des Datensatzes, Mediandaten oder Monte-Carlo simulierte Daten, abhängig. Die Werte der Typen BE und EF sind etwas besser. Eine Tendenz zu besseren Ergebnissen bei der Nutzung von Mediandaten ist nicht eindeutig erkennbar.

Die Güte der MLE/EM-Schätzungen ist von der Anzahl der zu schätzenden Unterverteilungen nahezu unabhängig. Bei den Mediandaten ist annähernd keine Abweichung vom idealen Verlauf vorhanden. Die Schätzungen bei den mit Monte-Carlo Methode erzeugten Datensätzen liegen im gleichen Bereich wie bei den Verteilungen mit zwei Unterverteilungen.

Bild 6.5 zeigt die besten Parameterschätzungen der Typen AB und EF.

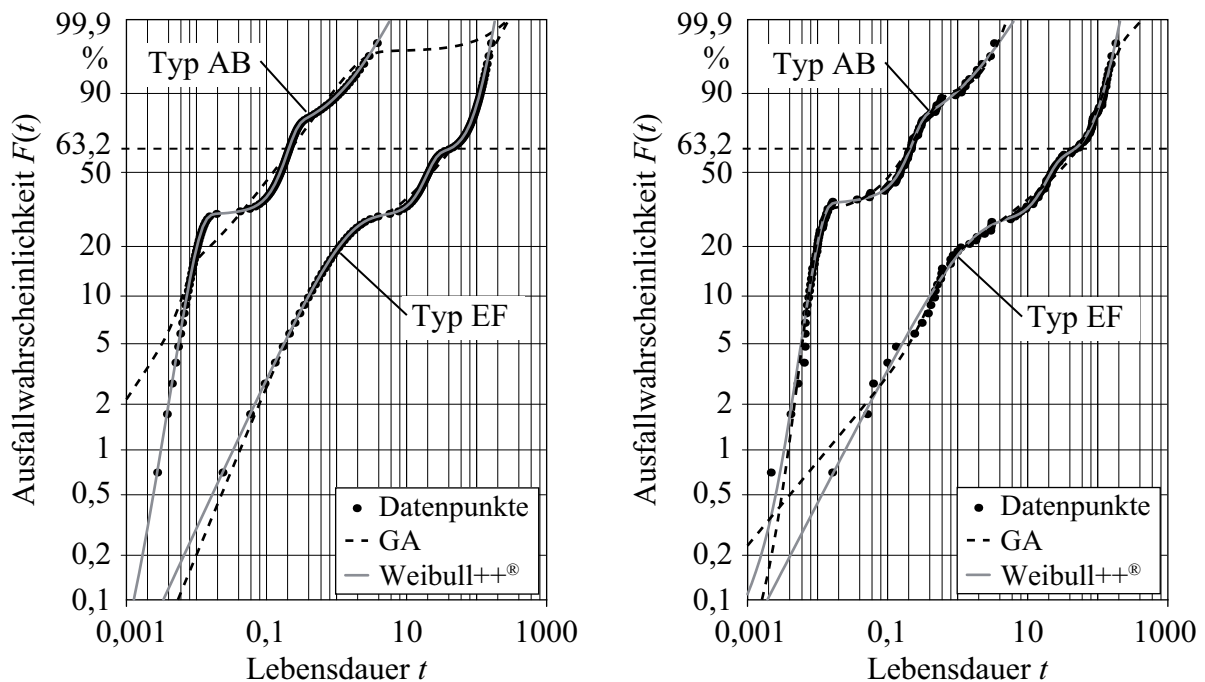


Bild 6.5: Geschätzte Verteilungsfunktionen für Typ AB und Typ EF mit Weibull++[®] und genetischem Algorithmus, Mediandaten (links), Daten über Monte-Carlo Methode (rechts)

Dargestellt ist der Verteilungstyp AB mit großen Abweichungen und der Verteilungstyp EF mit den kleinsten Abweichungen. Für beide Datentypen zeigen die mit MLE/EM-Algorithmus geschätzten Verteilungen einen idealen Verlauf. Die Verläufe der mittels genetischem Algorithmus geschätzten Verteilungen bilden die Datensätze für den Typ EF bei Mediandaten und bei beiden Typen für die Monte-Carlo simulierten Daten gut ab.

Deutlich schlechter ist die Anpassung für den Datensatz des Typs AB mit Mediandaten. Hier zeigt sich eine sehr schlechte Abbildung der Daten im Bereich des Sattelpunktes und bei hohen Ausfallwahrscheinlichkeiten. Die Verteilung der Monte-Carlo simulierten Daten bewirkt eine Verhinderung der Konvergenz zu dem lokalen Optimum, welches bei den Mediandaten dominant ist. Die großen Abweichungen der Lösungen von AB resultieren auch aus dem sehr groß gewählten Lösungsraum von T von $[0, 500]$. Durch eine Verkleinerung des Lösungsraumes von T auf $[0, 10]$ für Typ AB sind deutlich bessere Schätzungen möglich.

Die Schätzung von Verteilungen mit sehr weit auseinanderliegenden Werten von T ist problematisch. Bei Typ AF liegt eine Größenordnung von 10^4 zwischen kleinstem und größtem Wert. Durch die Unterschiede ergeben sich, bei gleichem Lösungsraum und gleicher Genauigkeit der Lösungen, stark unterschiedliche Wahrscheinlichkeiten, eine gute Schätzung zu erhalten. Dies soll anhand von folgendem Beispiel gezeigt werden:

Die zu schätzenden Verteilungen seien mit $b_a = 3$, $T_a = 0,01$ und $b_b = 3$, $T_b = 100$ und der Lösungsraum für $L_T = [0,001, 200]$ und die Genauigkeit der Lösungen mit $m_T = 4$ festgelegt. Der Formparameter der geschätzten Verteilung ist mit $b_g = 3$ konstant. Der Wert der charakteristischen Lebensdauer der geschätzten Verteilungen wird innerhalb des Intervalls L_T mit der Schrittweite 10^{-4} , resultierend aus m_T , verändert. Mit der zu schätzenden Verteilung und der geschätzten Verteilung wird über Gl. (6.3) die Abweichung bestimmt. Damit können die Parameterschätzungen identifiziert werden, die einen festgelegten Wert von D nicht überschreiten. Dadurch ergibt sich eine Anzahl an Parameterschätzungen n_{Va} bzw. n_{Vb} , die die erforderliche Güte D unterschreiten. Daraus lässt sich das Verhältnis σ berechnen.

$$\sigma = \frac{n_{Vb}}{n_{Va}} \quad (6.4)$$

In Bild 6.6 ist das Verhältnis σ der möglichen Parameterschätzungen beider Verteilungen über den Grenzwerten von D dargestellt. Bei einer zulässigen Abweichung D von 0,001 ergibt sich ein Verhältnis σ von etwa 2400. Das bedeutet, eine geringe Abweichung von der geschätzten Verteilung ist bei der Verteilung mit T_b 2400mal häufiger innerhalb des Lösungsraumes möglich als bei der Verteilung mit T_a . Mit zunehmendem D steigt das Verhältnis σ auf einen Wert von 10^4 , was genau dem Verhältnis T_b/T_a

entspricht. Der Anstieg im Bereich bis $D = 0,008$ und die Welligkeit sind auf die Schrittweite von 10^{-4} zwischen den einzelnen Werten von T zurückzuführen.

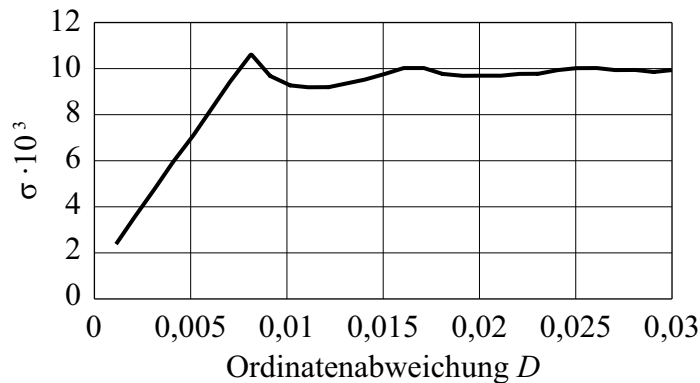


Bild 6.6: Verhältnis der Parameterschätzungen σ der Beispielverteilungen über dem Ordinatenabstand D

Dieses große Verhältnis von σ führt dazu, dass bei stark unterschiedlichen Verteilungen die Verteilung mit dem höheren Parameterwert genauer approximiert wird, was bei Mischverteilungen mit stark unterschiedlichen Unterverteilungen zu einer ungleichen Approximation der jeweiligen Unterverteilungen führt. Dies zeigt sich bei Typ AF, bei dem sich ein Verhältnis von der maximalen zur minimalen charakteristischen Lebensdauer von 10^4 ergibt. Durch die Reduktion der Größe des Lösungsraumes bei der Parameterschätzung von Typ AB erhöht sich die Wahrscheinlichkeit einer Annäherung an T durch die Vergrößerung der Lösungsdichte, siehe Gl. (4.27).

Zur weiteren Untersuchung der Güte der Schätzungen bei mehreren Unterverteilungen werden Mischverteilungen mit vier Unterverteilungen genutzt. Die Parameter des genetischen Algorithmus bleiben bis auf die Anzahl der Unterverteilungen gleich wie die bei den Mischverteilungen mit drei Unterverteilungen. Die Datensätze sind im Anhang in Tabelle A2.26 bis Tabelle A2.28 zu finden.

Die Ergebnisse der Parameterschätzungen mittels genetischem Algorithmus und Weibull++[®] sind in Bild 6.7 dargestellt. Bei den Ergebnissen der Schätzung von vier Unterverteilungen zeigt sich tendenziell dasselbe Bild wie bei der Schätzung von drei Unterverteilungen. Die Abweichungen sind bei Typ ADF am größten, was sich durch das große Verhältnis von maximaler zu minimaler charakteristischer Lebensdauer erklären lässt. Das Niveau der Abweichung ist etwas geringer, da durch die zusätzlich vorhandene Unterverteilung eine genauere Abbildung der Datenpunkte möglich ist. Die beste Schätzung ist die des Typs CDE mit einem Verlauf ohne scharfe Knickstellen und Sattelpunkte. Die Güte der Schätzung des Typs BDE liegt zwischen den Schätzungen der beiden anderen Typen.

Die Schätzungen mittels MLE/EM-Algorithmus zeigen auch bei diesen Verteilungen sehr gute Ergebnisse und nur eine sehr geringe Abhängigkeit von der Art des zu schätzenden Datensatzes.

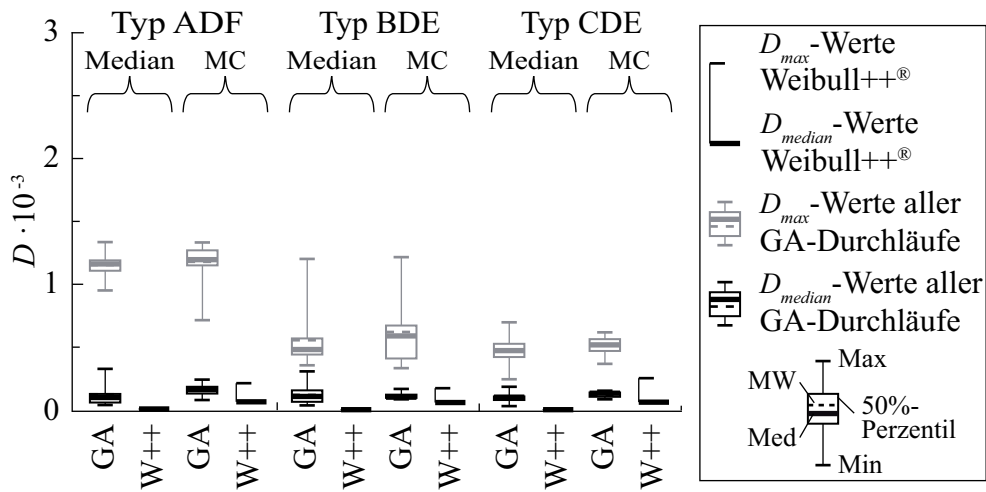


Bild 6.7: Ergebnisse der Parameterschätzungen der Mischverteilungen mit vier Unterverteilungen mit Weibull++[®] und genetischem Algorithmus

Die Verläufe der Mischverteilungen von Typ ADF und CDE für die unterschiedlichen Datensätze sind in Bild 6.8 dargestellt. Für den genetischen Algorithmus sind die besten Schätzungen aus 20 Replikationen gezeigt.

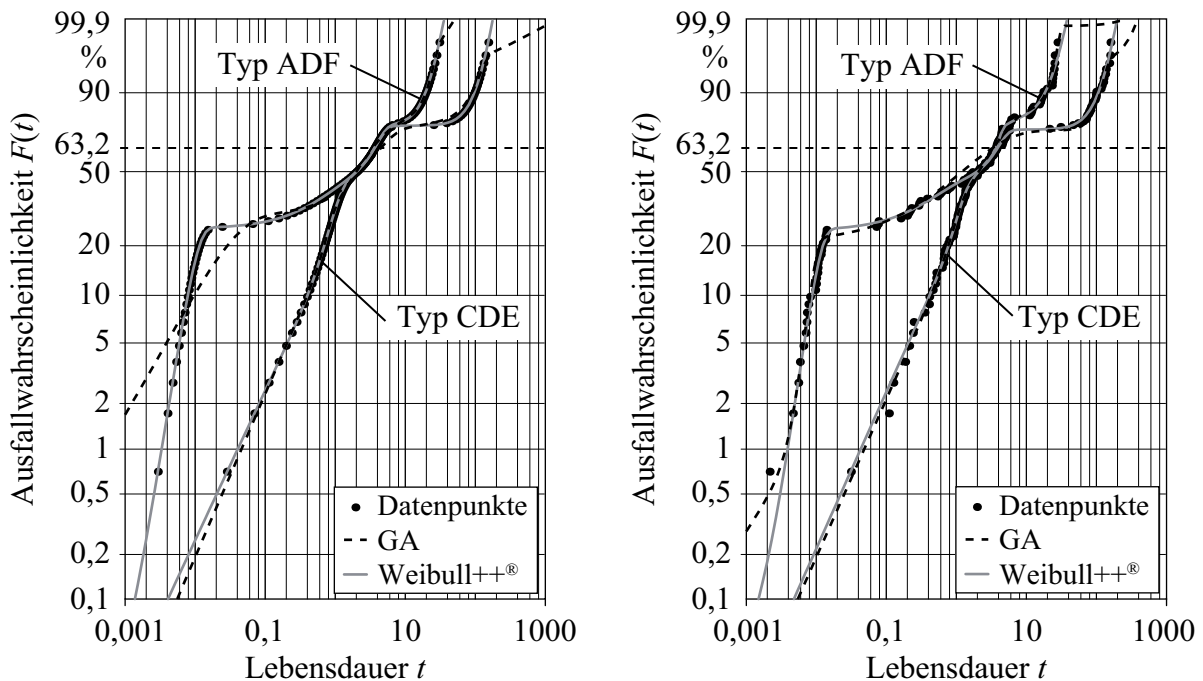


Bild 6.8: Geschätzte Verteilungsfunktionen für Typ ADF und Typ CDE mit Weibull++[®] und genetischem Algorithmus, Mediandaten (links), Daten über Monte-Carlo Methode (rechts)

Wie schon bei den Mischverteilungen mit drei Unterverteilungen zeigt sich auch bei der Verteilung ADF die Schwierigkeit der Abbildung der Daten im Bereich des Sattelpunktes. Zusätzlich ist eine Abweichung aller Verläufe vom idealen Verlauf bei hohen Ausfallwahrscheinlichkeiten zu erkennen. Insgesamt werden die Datenpunkte bis auf Typ ADF mit Mediandaten gut durch die geschätzten Verteilungen abgebildet.

Bei der Schätzung von Mischverteilungen mit mehreren Unterverteilungen zeigt sich, dass der MLE/EM-Algorithmus sehr gute Ergebnisse liefert, die vom genetischen Algorithmus nicht immer erreicht werden können. Problematisch sind vor allem scharfe Knickstellen und Mischverteilungen mit sehr unterschiedlichen Parametern. Zusammenfassend kann der genetische Algorithmus in den meisten Fällen gute Parameterschätzungen liefern.

6.3.3 Parameterschätzung der abgeleiteten Mischverteilungen mit t_0

In diesem Abschnitt wird der genetische Algorithmus zur Parameterschätzung auf die in Kapitel 5.1.3 vorgestellten Mischverteilungen angewandt. Bei diesen Mischverteilungen ist jeweils ein zusätzlicher Parameter, die ausfallfreie Zeit t_0 , zu schätzen. Dadurch erhöht sich die Anzahl der zu schätzenden Parameter auf sieben. Die Datensätze zu den Verteilungen sind im Anhang in Tabelle A2.29 bis Tabelle A2.37 zu finden, die Parameter in Tabelle A1.11. Die Ergebnisse der Parameterschätzungen der Mischverteilungen mit ausfallfreier Zeit sind in Bild 6.9 dargestellt.

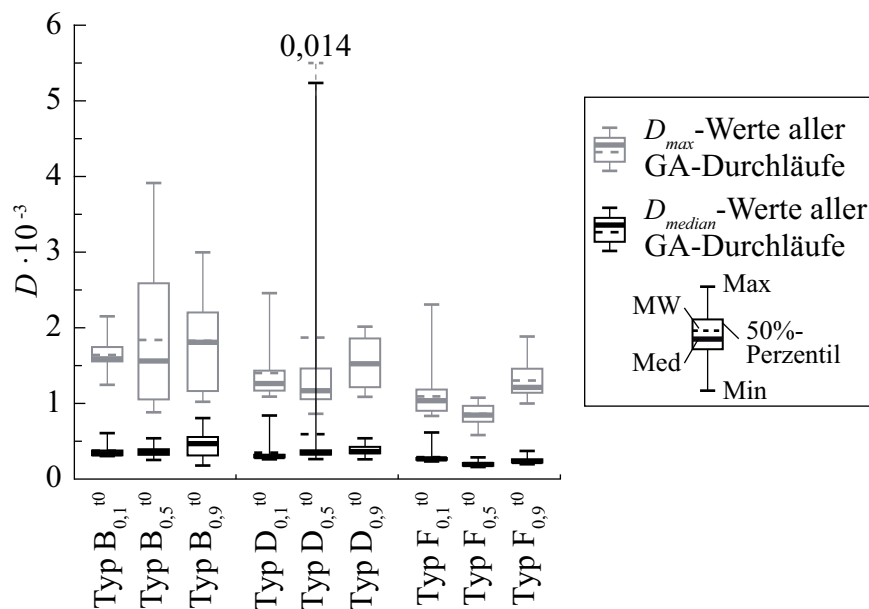


Bild 6.9: Ergebnisse der Parameterschätzungen der Mischverteilungen nach Jiang und Kececioglu erweitert um t_0 mit genetischem Algorithmus

Ein Vergleichsverfahren zum genetischen Algorithmus ist für die Weibull-Mischverteilungen mit ausfallfreier Zeit nicht vorhanden, da es in keiner bekannten

kommerziellen Software implementiert ist. Brodbeck hatte in [135] Parameterschätzungen mittels MLE/EM-Algorithmus durchgeführt und auf Schwierigkeiten im Zusammenhang mit der Schätzung der ausfallfreien Zeit hingewiesen.

Durch die Einführung des zusätzlichen Parameters zeigen alle Parameterschätzungen größere Werte von D_{max} und D_{median} als bei den Varianten ohne ausfallfreie Zeit. Die deutlichste Verschlechterung der Ordinatenabstände ergibt sich bei Typ B. Ausschlaggebend dafür ist der scharfe Knick im Verlauf der Verteilung, der sich bei großen Werten von f_{tb} ergibt, siehe Bild 6.10.

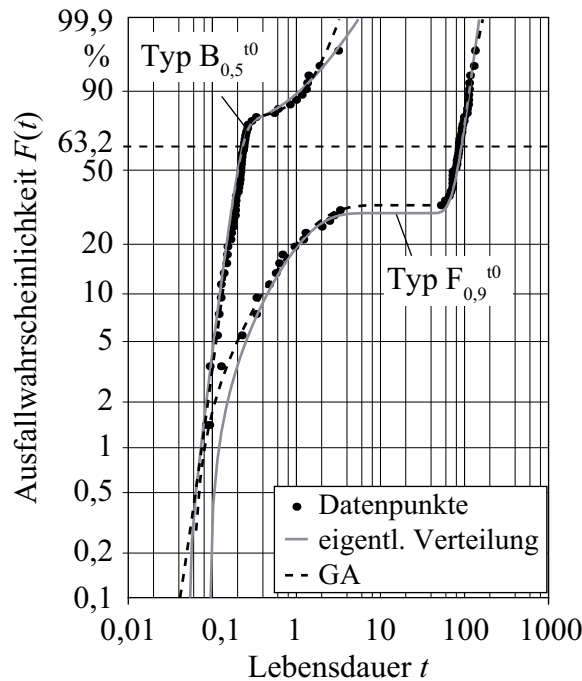


Bild 6.10: Geschätzte Verteilungsfunktionen für Typ $B_{0,5}^{t_0}$ und Typ $F_{0,9}^{t_0}$ mit genetischem Algorithmus

Der Verläufe von Mischverteilungstyp $B_{0,5}^{t_0}$ mit großem Ordinatenabstand und der Verlauf von Typ $F_{0,9}^{t_0}$ mit geringem Ordinatenabstand sind dargestellt. Für beide Mischverteilungstypen sind sowohl der Verlauf der ursprünglichen Verteilungskurve als auch die mittels genetischem Algorithmus geschätzte Verteilung dargestellt. Beide Verläufe der geschätzten Verteilungen bilden die Datenpunkte gut ab. Die großen Ordinatenabstände sind zum Teil auch auf die teilweise starke Streuung in den Monte-Carlo simulierten Datensätzen zurückzuführen.

Durch die Hinzunahme der ausfallfreien Zeit als dritten Parameter der Unterverteilungen erhöhen sich hauptsächlich die Werte von D_{max} deutlich. Parameterschätzungen mit guter Abbildung der Datenpunkte sind möglich. Die durch große ausfallfreie Zeiten verursachten Knickstellen können bei der Parameterschätzung zu großen Abweichungen bei den Ordinatenabständen führen. Generell sind die Parameterschätzungen nicht deutlich schlechter als bei der Schätzung der Verteilungen ohne ausfallfreie Zeit.

6.4 Parameterschätzung von Systemen

Der zweite Typ der gemischten Verteilungen sind neben den Mischverteilungen die booleschen Systeme. Die Verteilungen, die in diesem Abschnitt untersucht werden, sind in Kapitel 5.2 beschrieben. Die Parameterschätzung von booleschen Systemen unterscheidet sich gegenüber Mischverteilungen deutlich durch das Fehlen der Gewichtungsfaktoren p_i . Zudem unterscheidet sich die Berechnung der Fitnessfunktion für die Systeme über das Produkt von der Berechnung über die Summe bei Mischverteilungen, siehe Kapitel 4.3. Eine weitere Unterscheidung bei der Fitnessberechnung ist die Unterteilung in Serien- und Parallelsysteme, siehe Gl. (4.14) und Gl. (4.15).

Im ersten Teil des Kapitels werden die Parameterschätzungen der in Kapitel 5.2.2 vorgestellten Seriensysteme mit zwei Unterverteilungen vorgestellt. Der zweite Teil des Kapitels beschäftigt sich mit der Schätzung der Parallelsysteme aus Kapitel 5.2.3. Die jeweiligen Datensätze zu den Verteilungen sind im Anhang A2 zu finden.

6.4.1 Parameterschätzung von Seriensystemen

Die Parameterschätzung von Seriensystemen wurde mit den im Anhang angegebenen Parametern, siehe Tabelle A1.12, durchgeführt. Die Daten wurden mit der Monte-Carlo Methode erzeugt und sind im Anhang in Tabelle A2.38 bis Tabelle A2.43 zu finden. Für die Schätzung von Seriensystemen ist keine kommerzielle Software bekannt, die die Parameterschätzung von booleschen Systemen unterstützt, weshalb keine Vergleichsmethode zur Verfügung steht.

Die Ergebnisse der Schätzungen der Seriensystemtypen sind in Bild 6.11 dargestellt. Die Grundlage für die Box-Plots sind 20 Replikationen des genetischen Algorithmus. Aufgetragen sind für jede Verteilung die Box-Plots für D_{max} und D_{median} .

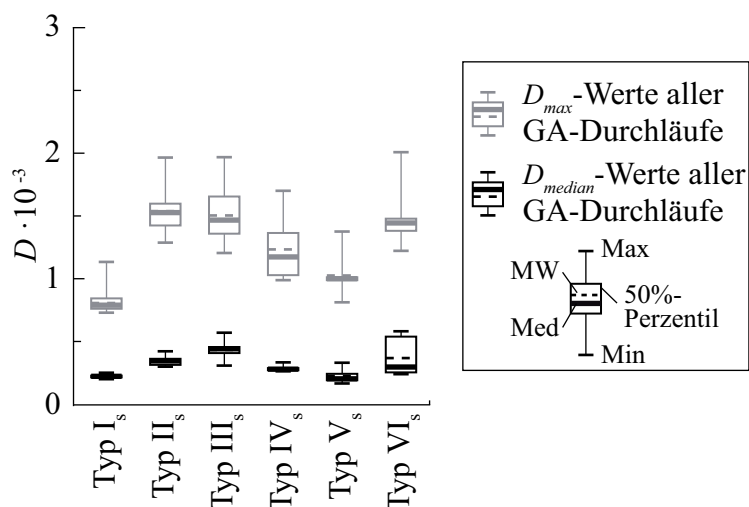


Bild 6.11: Ergebnisse der Parameterschätzungen der Seriensysteme mit genetischem Algorithmus

Die Ordinatenabstände der geschätzten Systeme sind mit denen der Mischverteilungen mit zwei Unterverteilungen bei gleicher Stichprobengröße $n = 50$ vergleichbar. Es ist eine Tendenz zu schlechteren Parameterschätzungen bei den Typen mit Schnittpunkt bei geringeren Ausfallwahrscheinlichkeiten erkennbar. Dies resultiert aus dem ausgeprägteren Übergangsbereich zwischen den Verteilungen, der für Abweichungen der Monte-Carlo simulierten Daten anfälliger ist, siehe Bild 6.12. Dargestellt sind die geschätzten Verteilungen der 20 Replikationen für Seriensysteme des Typs I_s und VI_s .

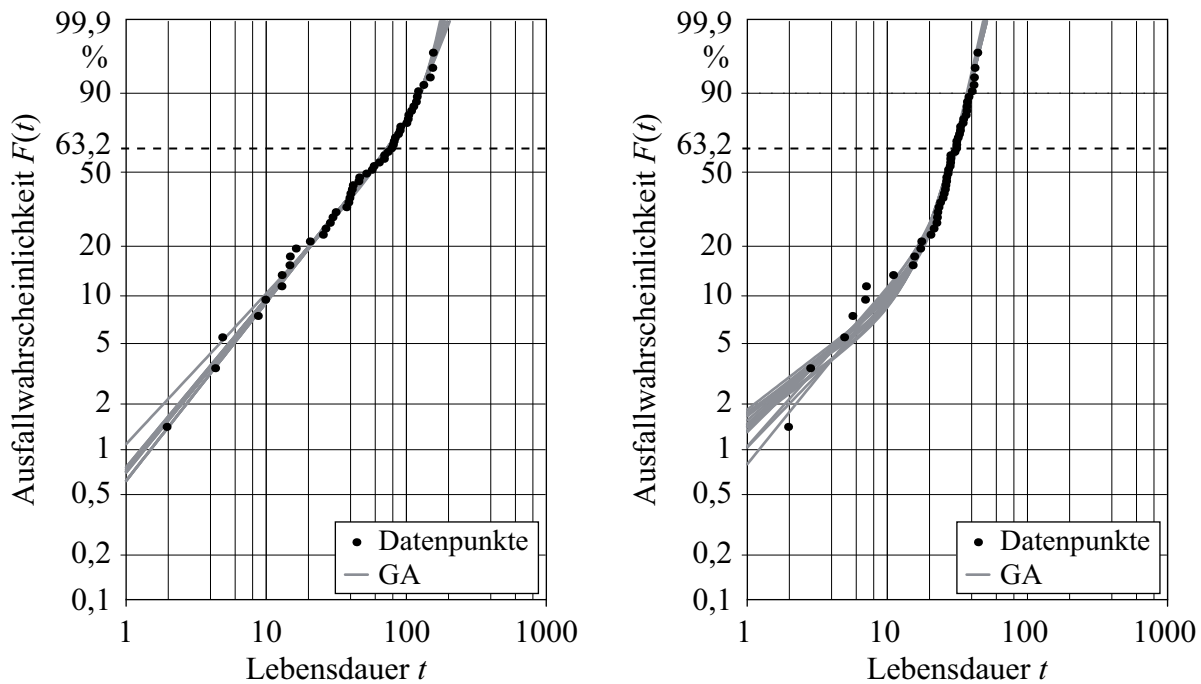


Bild 6.12: Geschätzte Verteilungsfunktionen der Seriensysteme Typ I_s (links) und Typ VI_s (rechts)

Gut erkennbar ist die sehr gute Abbildung der Daten bei Typ I_s . Die Streuung des Datensatzes ist nicht sehr ausgeprägt. Die Verteilungen des Typs VI_s zeigen eine größere Streuung im Bereich geringer Ausfallwahrscheinlichkeiten, bedingt durch die große Streuung der Daten in diesem Bereich. Die Abbildung der Datenpunkte im Bereich größer 20 % Ausfallwahrscheinlichkeit ist sehr gut.

Zur Untersuchung der Seriensysteme mit ausfallfreier Zeit werden die in Kapitel 5.2.2 vorgestellten, aus den Typen II_s und V_s abgeleiteten Systeme genutzt. Die Parameterschätzung der Seriensysteme mit ausfallfreier Zeit wurde mit den in Tabelle A1.13 gezeigten Parametern durchgeführt. Die Datensätze sind in Tabelle A2.44 bis Tabelle A2.47 im Anhang A2 zu finden. Die Ergebnisse der Parameterschätzungen der Datensätze zeigt Bild 6.13 links.

Durch die zwei weiteren zu schätzenden Parameter vergrößern sich sowohl der Betrag der Werte von D_{max} und D_{median} als auch die Streuungen. Besonders stark ist der Anstieg bei Typ $II_{s, 0,1}^{10}$ und Typ $V_{s, 0,1}^{10}$. Dies ist auf die sehr geringen Werte der

charakteristischen Lebensdauern von $T_1 = 1,1$ und $T_2 = 3,8$ für Typ $II_{s,0,1}^{10}$ und $T_1 = 1,1$ und $T_2 = 4,8$ für Typ $V_{s,0,1}^{10}$ zurückzuführen, die nur unzureichend angenähert werden können. Dies ist ebenfalls an den großen Streuungen der geschätzten Verteilungen im Bereich geringer Ausfallwahrscheinlichkeiten in Bild 6.13 rechts erkennbar. In diesem Bereich wirken sich auch die Streuungen der Monte-Carlo simulierten Daten besonders stark aus. Zudem wird der Bereich der geringen Ausfallwahrscheinlichkeiten nur durch wenige Punkte unterhalb des Schnittpunktes definiert. Dadurch ist, trotz Gewichtung der Fitnessfunktion, die Anpassung in den Bereichen höherer Ausfallwahrscheinlichkeit mit mehr Datenpunkten stärker ausschlaggebend. Für größere Werte von t_0 bei den Typen $II_{s,0,5}^{10}$ und $V_{s,0,5}^{10}$ sind die Medianwerte und die 50 %-Perzentile nur unwesentlich größer als bei den Systemen ohne ausfallfreie Zeit.

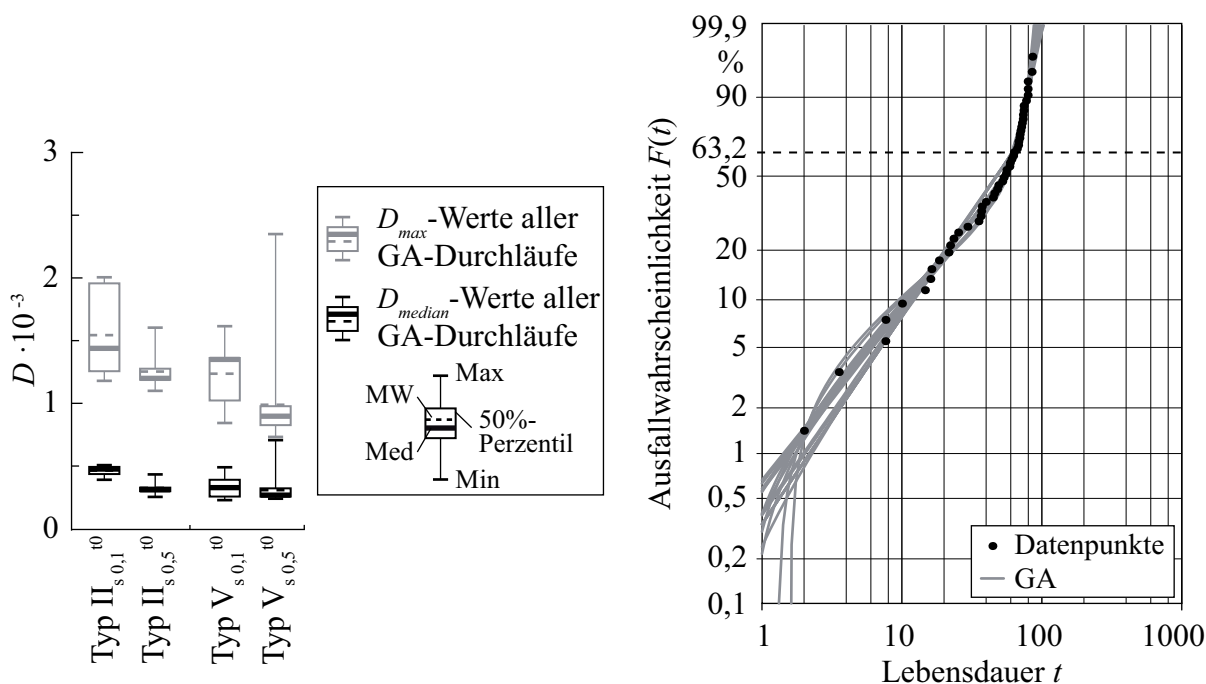


Bild 6.13: Ergebnisse der Parameterschätzungen der Seriensysteme mit ausfallfreier Zeit t_0 mit genetischem Algorithmus (links) und Schätzungen von Typ $V_{s,0,1}^{10}$ (rechts)

Die Parameterschätzung von booleschen Seriensystemen ist mit dem genetischen Algorithmus gut möglich. Für die gezeigten Beispielsysteme sind gute bis sehr gute Schätzungen der Verteilungsparameter erreicht worden. Es zeigen sich ähnliche Schätzergebnisse wie bei den Mischverteilungen. Der Vorteil des Wegfalls der zusätzlich zu schätzenden Gewichtungsfaktoren p_i fällt geringer aus als erwartet. Die Schätzung von Systemen mit Systemelementen mit ausfallfreier Zeit ist ähnlich wie bei den Mischverteilungen mit einer höheren Streuung verbunden. Hier wirkt sich wie bei den Mischverteilungen die Vergrößerung der Anzahl an möglichen Lösungsvarianten durch einen weiteren Parameter t_0 negativ aus.

6.4.2 Parameterschätzung von Parallelsystemen

In diesem Abschnitt wird die zweite Variante der booleschen Systeme, werden die Parallelsysteme, untersucht. Grundlage sind die in Kapitel 5.2.3 vorgestellten Parallelsysteme. Die zugehörigen Datensätze sind in Tabelle A2.48 bis Tabelle A2.53 im Anhang A2 aufgeführt. Für die Parameterschätzung werden dieselben Parameter verwendet, die bereits für die Schätzung der Seriensysteme verwendet wurden, siehe Anhang Tabelle A1.12.

Die Box-Plots der Ergebnisse der Parameterschätzungen der Parallelsysteme Typ I_p bis Typ VI_p zeigt Bild 6.14.

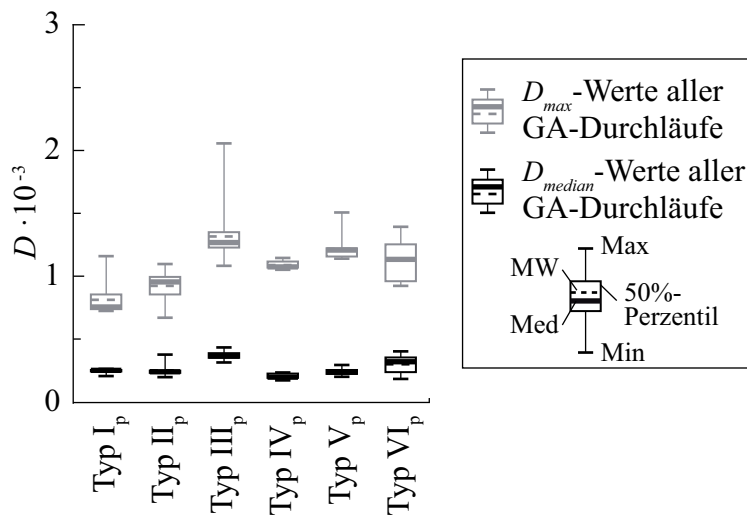


Bild 6.14: Ergebnisse der Parameterschätzungen der Parallelsysteme mit genetischem Algorithmus

Die Ergebnisse der Parameterschätzungen der Parallelsysteme sind mit denen der Seriensysteme vergleichbar, die Streuungen und auch die Beträge sind etwas geringer. Wie bei den Seriensystemen ist eine Tendenz zu schlechteren Parameterschätzungen bei den Parallelsystemen Typ III_p und Typ VI_p erkennbar. Hier wirkt sich ebenfalls die geringere Anzahl an Ausfallpunkten unterhalb des Schnittbereichs negativ auf die Anpassung an die Datenpunkte aus. Auf die Darstellung der geschätzten Parallelsysteme im Weibullpapier wird, aufgrund der nur geringen Abweichungen von den Datenpunkten bei allen 20 Replikationen, verzichtet.

Zur Untersuchung der Parallelsysteme mit ausfallfreier Zeit werden die in Kapitel 5.2.3 vorgestellten, aus den Typen II_p und V_p abgeleiteten Systeme genutzt. Die daraus über die Monte-Carlo Methode generierten Daten finden sich im Anhang in Tabelle A2.54 bis Tabelle A2.57. Wie bei den Parallelsystemen ohne ausfallfreie Zeit werden für die Systeme mit ausfallfreier Zeit dieselben Parameter des genetischen Algorithmus wie bei den Seriensystemen mit ausfallfreier Zeit verwendet, siehe Tabelle A1.13.

Die Ergebnisse der Schätzung der Parallelsysteme mit ausfallfreier Zeit sind in Bild 6.15 links gezeigt. Die ausfallfreien Zeiten in den Parallelsystemen bewirken eine Erhöhung der Werte von D_{max} , die Werte für D_{median} bleiben annähernd gleich wie bei den Parallelsystemen ohne ausfallfreie Zeit. Die Erhöhung der maximalen Ordinatenabstände wird vor allem durch die schlechtere Anpassung im Bereich geringer und hoher Ausfallwahrscheinlichkeiten verursacht. Im rechten Bild dargestellt sind alle 20 Replikationen der Schätzung von Typ $II_{p,0,5}^{t_0}$. Dies sind die Schätzungen mit den größten Medianwerten der Ordinatenabweichungen D_{max} und D_{median} . Die Ergebnisse sind, wie der Box-Plot erwarten lässt, alle sehr ähnlich und zeigen nur im niedrigen und hohen Ausfallwahrscheinlichkeitsbereich Streuungen. Der mittlere Bereich der Datenpunkte wird von allen Schätzungen sehr gut wiedergegeben.

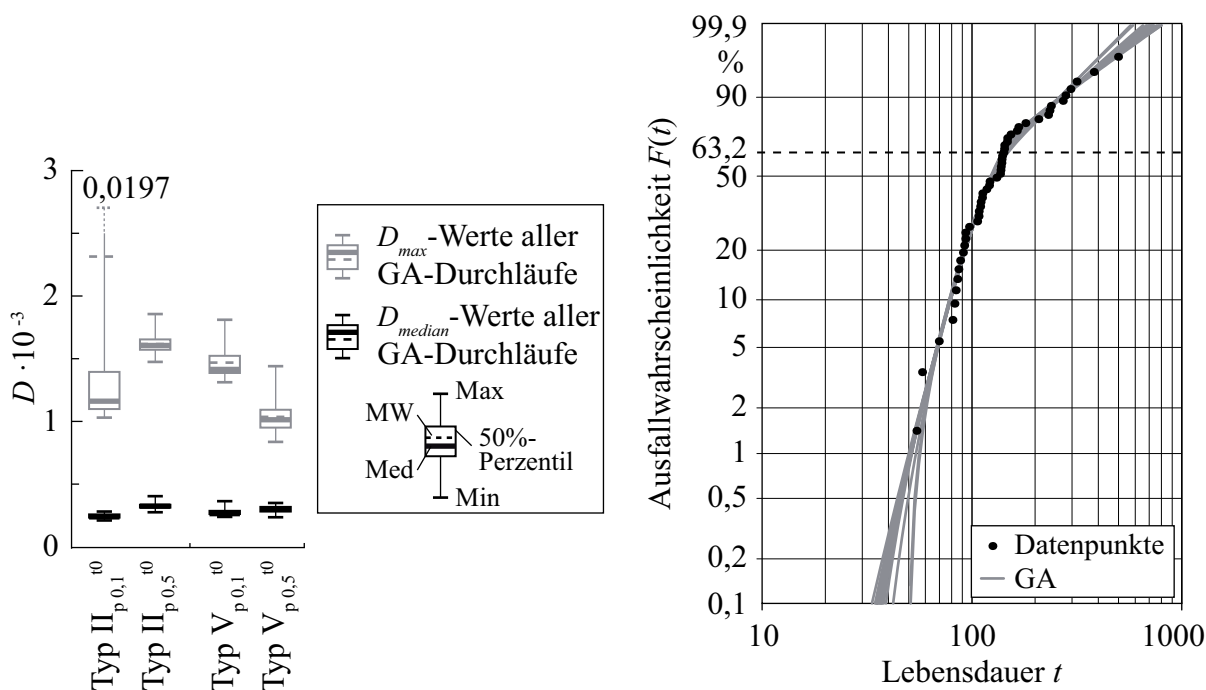


Bild 6.15: Ergebnisse der Parameterschätzungen der Parallelsysteme mit ausfallfreier Zeit t_0 mit genetischem Algorithmus (links) und Schätzungen von $Typ II_{p,0,5}^{t_0}$ (rechts)

Parameterschätzungen boolescher Parallelsysteme sind mit dem genetischen Algorithmus gut möglich. Wie schon bei den Seriensystemen sind für die gezeigten Beispiele gute bis sehr gute Schätzungen von Verteilungsparametern möglich. Die Ergebnisse der Schätzungen sind beim Vergleich der maximalen Ordinatenabstände und dem Median der Ordinatenabstände mit denen der Seriensysteme und Mischverteilungen vergleichbar. Parallelsysteme mit ausfallfreier Zeit erhöhen bei den Schätzungen die Abweichungen der Ordinatenwerte bei den gezeigten Beispielen, jedoch nicht die Streuung. Die Ergebnisse bilden die Datenpunkte gut ab.

7 Schätzung der Unterverteilungsanzahl und Bewertung der Anwendung

Neben der Schätzung von Parametern von Unterverteilungen oder Systemkomponenten ist es mit dem genetischen Algorithmus möglich, die Anzahl der Unterverteilungen einer Mischverteilung oder eines Systems zu schätzen. Bei analytischen Methoden wie MLE/EM ist es erforderlich, schon zu Beginn die Zahl der Unterverteilungen festzulegen. In diesem Kapitel wird die Möglichkeit zur Schätzung der Anzahl an Unterverteilungen mittels des genetischen Algorithmus aufgezeigt und die erzielten Schätzungen diskutiert. Zuletzt wird die Einsetzbarkeit des genetischen Algorithmus als Methode zur Schätzung von Verteilungsparametern von Mischverteilungen und booleschen Systemen bewertet.

7.1 Schätzung der Unterverteilungsanzahl

Gegenüber analytischen Methoden bietet der genetische Algorithmus den Vorteil, nicht nur für eine festgelegte Anzahl an Verteilungsparametern Schätzungen durchführen zu können. Dies ist für die bekannten analytischen Methoden, zur Aufstellung der Gleichungen beispielsweise der Likelihoodfunktion, erforderlich. Beim genetischen Algorithmus kann die Anzahl der zu schätzenden Unterverteilungen über ein Intervall vorgegeben werden. Der Ablauf der Parameterschätzungen ist dabei prinzipiell derselbe.

Es werden zuerst zufällig Eltern der ersten Generation erzeugt. Die Anzahl der Unterverteilungen der Eltern wird wie bei allen anderen Parametern zufällig gewählt. Dadurch ergibt sich für die erste Generation eine zufällige Verteilung der Unterverteilungsanzahl. Im Verlauf des Algorithmus ergibt sich für die Mutation keine Änderung des Ablaufes. Bei der Rekombination muss berücksichtigt werden, dass nur Individuen mit gleicher Anzahl an Unterverteilungen miteinander rekombiniert werden können. Dazu muss bei der Paarungsselektion die jeweilige Anzahl an Unterverteilungen als Kriterium berücksichtigt werden. Zu jedem gewählten Individuum mit einer bestimmten Anzahl Unterverteilungen muss mindestens ein Individuum mit gleicher Anzahl an Unterverteilungen vorhanden sein. Für die weiteren Ablaufschritte der Fitnessberechnung und der Populationsselektion ist die Anzahl an Unterverteilungen der Individuen unerheblich. Bei der Umweltselektion ist es wiederum erforderlich, mindestens zwei Individuen mit derselben Anzahl an Unterverteilungen in der Population zu haben, um eine weitere Verbesserung der Individuen über die Rekombination zu ermöglichen. In

diesem Zusammenhang ist es wichtig, eine Selektionsvariante für die Umweltselektion zu wählen, die bessere Individuen bei der Auswahl bevorzugt. Wäre dies nicht der Fall, wäre es möglich, dass bestimmte Unterverteilungszahlen aus der Generation ausgeschlossen werden, die eine hohe Fitness und ein Potential für noch bessere Fitnesswerte haben.

Für den Mischverteilungstyp E und das Seriensystem Typ II_s werden Parameterschätzungen mit dem Intervall der Unterverteilungen von [1, 10] durchgeführt, siehe Tabelle A1.14 und Tabelle A1.15. Für jede Schätzung werden je 100 Replikationen durchgeführt. Es werden die Monte-Carlo simulierten Datensätze aus Tabelle A2.20 für die Mischverteilung und die Werte aus Tabelle A2.39 für das Seriensystem verwendet. Durch die Monte-Carlo simulierten Datensätze konnten stochastische Effekte einbezogen werden, die die Bestimmung der Unterverteilungsanzahl beeinflussen können.

7.1.1 Schätzung der Unterverteilungen in Mischverteilungen

In Bild 7.1 ist die geschätzte Anzahl an Unterverteilungen des besten Individuums für die Mischverteilung Typ E für 100 Replikationen dargestellt.

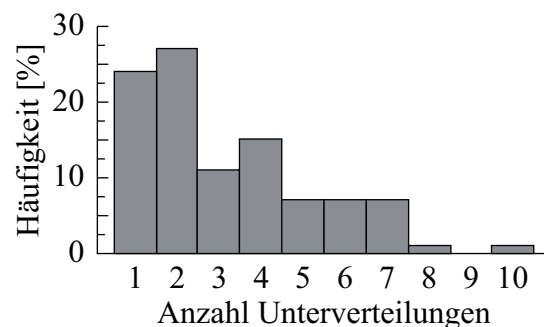


Bild 7.1: Verteilung der Anzahl der Unterverteilungen bei den besten Individuen und 100 Replikationen bei der Schätzung der Mischverteilung Typ E

Bei der Schätzung der Unterverteilungsanzahl der Mischverteilung ist eine rechtschiefe Verteilung zu erkennen. Das Intervall der Unterverteilungen wird bis auf die Anzahl von neun Unterverteilungen komplett abgedeckt. Es werden jedoch häufiger Verteilungen mit wenigen Unterverteilungen geschätzt als mit vielen. Am häufigsten treten Schätzungen mit zwei Unterverteilungen, der richtigen Anzahl auf. Kaum seltener sind die Schätzungen mit nur einer Unterverteilung.

Anhand von Bild 7.2 kann die Ursache der Schätzungen der Unterverteilungsanzahl erklärt werden. Im linken Bild dargestellt ist eine Schätzung mit zehn Unterverteilungen. Eine der Unterverteilungen liegt mit $p = 0,00084$ außerhalb des dargestellten Bereichs. Die Anpassung der Mischverteilung an die Datenpunkte ist nur im Bereich bis etwa 40 % Ausfallwahrscheinlichkeit ausreichend. Im darüber liegenden Bereich ist keine Anpassung mehr vorhanden. Gut zu erkennen ist die Dominanz einiger weniger

Unterverteilungen mit hohen Gewichtungen, die den Verlauf der Mischverteilung maßgeblich bestimmen. Es sind sechs Verteilungen vorhanden, die einen Anteil an der Gesamtverteilung von jeweils 1 % oder weniger haben. Es treten somit in der geschätzten Verteilung Unterverteilungen auf, die keinen merklichen Einfluss haben. Punktuell können diese aber zu einer besseren Fitness führen, wodurch sie bei der Selektion bevorzugt ausgewählt werden. Ein Großteil der Unterverteilungen ist für die Approximation des Datensatzes nicht notwendig.

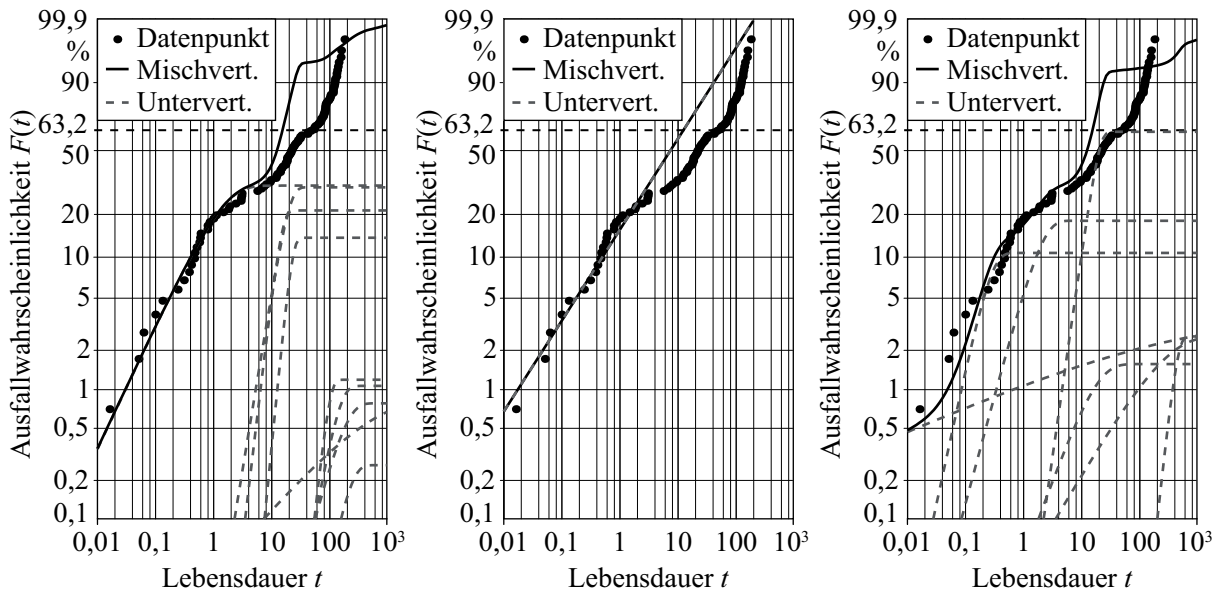


Bild 7.2: Geschätzte Verteilungen des Mischverteilungstyps E mit Unterverteilungen, mit zehn Unterverteilungen (links), mit einer Unterverteilung (Mitte) und die Schätzung mit der geringsten Fitness mit sieben Unterverteilungen (rechts)

Im mittleren Bild ist bei der Schätzung nur eine Verteilung vorhanden, die den Bereich bis ungefähr 25 % Ausfallwahrscheinlichkeit gut, den Bereich darüber nicht mehr ausreichend abbildet. Dies kann auftreten, wenn die Einzelverteilungen in frühen Generationen eine weit überlegene Fitness besitzen und deshalb schon früh Mischverteilungen mit mehreren Unterverteilungen aus der Population ausscheiden.

In der rechten Darstellung in Bild 7.2 ist die Schätzung mit der geringsten Fitness aller 100 Replikationen gezeigt. Es sind sieben Unterverteilungen vorhanden, von denen drei für den Verlauf der Mischverteilung durch die hohen Gewichtungen maßgeblich sind. Die restlichen Verteilungen haben nur einen geringen oder gar keinen Einfluss auf den Gesamtverlauf. Bei der Unterverteilung mit etwa 60 % Gewichtung ist ein weiteres Phänomen erkennbar. Hier werden Welligkeiten, die durch die Monte-Carlo Methode im Verlauf der Datenpunkte entstehen können, approximiert. Diese sind für die Mischverteilung nicht relevant, können aber die Fitness der Gesamtschätzung verbessern.

Bei der Schätzung von Datensätzen aus Mediandaten treten dieselben Phänomene auf. Es ist daher nur eine geringe Abhängigkeit vom Datensatz erkennbar, soweit dieser nicht einen Verlauf hat, der durch Streuungen der Datenpunkte sehr stark von der eigentlichen Mischverteilung abweicht.

7.1.2 Schätzung der Unterverteilungen in Seriensystemen

Bei der Schätzung des Seriensystems sind die Häufigkeiten der Unterverteilungsanzahl gleichmäßiger verteilt als bei der Mischverteilung, siehe Bild 7.3. Schätzungen mit geringer Anzahl an Unterverteilungen treten seltener auf als bei Mischverteilungen. Ein System mit nur einer Unterverteilung tritt am Ende keines Durchlaufes als beste Lösung auf. Die richtige Anzahl an Unterverteilungen von $N = 2$ wird nur in 2 % der Fälle geschätzt. Die Häufigkeit der geschätzten Anzahl im Bereich zwischen vier und zehn Unterverteilungen ist nahezu gleich.

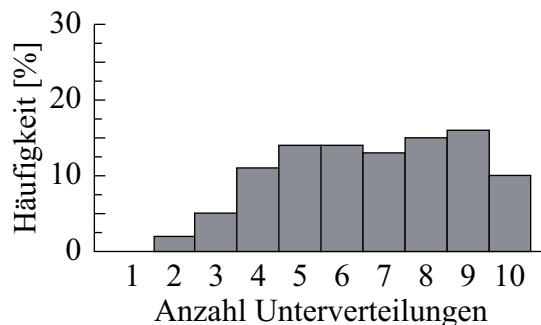


Bild 7.3: Verteilung der Anzahl der Unterverteilungen bei den besten Individuen und 100 Replikationen bei der Schätzung des Seriensystems Typ II_s

Das häufige Auftreten der Schätzungen mit vielen Unterverteilungen ist durch die Berechnung der Systemkurve erklärbar. Verteilungen, die eine hohe charakteristische Lebensdauer besitzen, haben kaum einen Einfluss auf den Verlauf der Systemkurve. In Bild 7.4 ist dieser Fall in der linken Darstellung gezeigt.

Es handelt sich hierbei um ein System mit zehn Unterverteilungen. Bei den meisten Unterverteilungen liegen die charakteristischen Lebensdauern weit höher als die beiden Verteilungen mit den geringsten charakteristischen Lebensdauern. Diese Verteilungen haben auf den Verlauf der Systemkurve keinen Einfluss. Die Systemkurve wird von den beiden am weitesten links liegenden Verteilungen bestimmt. Für die Fitnessberechnung sind im Prinzip nur zwei Verteilungen relevant. Die Selektion dieser Variante mit zehn Unterverteilungen beruht daher nicht auf der Vielzahl an Unterverteilungen, sondern nur auf der Güte der beiden maßgeblichen Verteilungen.

In der mittleren Darstellung in Bild 7.4 ist ein Seriensystem dargestellt, das nur aus zwei Unterverteilungen besteht, es sind keine unnötigen Unterverteilungen vorhanden. Im rechten Bild ist das System mit dem besten Fitnesswert dargestellt. Hierbei handelt

es sich um ein System aus neun Unterverteilungen. Von den dargestellten neun Unterverteilungen sind die drei links liegenden Verteilungen für den Verlauf der Systemkurve maßgebend. Die bessere Fitness dieser Variante beruht auf der besseren Abbildung der streuenden Datenpunkte im Bereich unterhalb von 40 % Ausfallwahrscheinlichkeit.

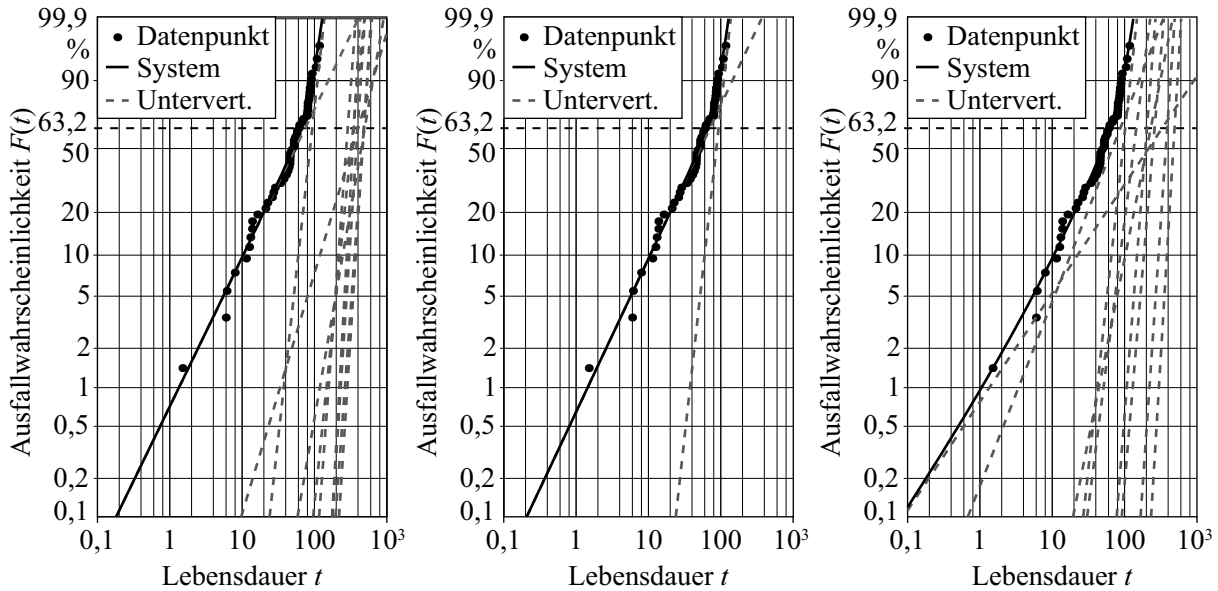


Bild 7.4: Geschätzte Verteilungen des Seriensystems II_s mit Unterverteilungen, mit zehn Unterverteilungen (links), mit zwei Unterverteilungen (Mitte) und die Schätzung mit der geringsten Fitness mit neun Unterverteilungen (rechts)

Verteilungen mit weniger als der zu schätzenden Anzahl an Unterverteilungen treten bei den gut separierbaren Verteilungen nicht auf. Die in diesen Verteilungen vorhandenen scharfen Knickstellen lassen sich nicht oder nur schlecht mit weniger Verteilungen darstellen. Bei der Schätzung von zu vielen Unterverteilungen steigt mit der Anzahl der Unterverteilungen auch die Wahrscheinlichkeit, dass Unterverteilungen vorhanden sind, die den Datensatz sehr gut abbilden. Die überzähligen Verteilungen haben aber keinen negativen Einfluss auf den Fitnesswert des Systems.

Bild 7.5 zeigt die Anzahl der geschätzten Unterverteilungen der einzelnen Individuen einer Generation und den Verlauf der Fitness über die Generationen hinweg. Gut erkennbar an dem Verlauf der Anzahl der Unterverteilungen ist das Verschwinden der Systeme mit nur einer und zwei Unterverteilungen nach nur zwei Generationen. Ab der 50sten Generation scheiden alle Lösungen mit mehr als sieben Unterverteilungen aus der Population aus, und ab der 80sten Generation sind nur noch Individuen mit einer Unterverteilungsanzahl von sieben vorhanden. Gut zu erkennen ist auch der relativ konstante Verlauf des Medians der Unterverteilungsanzahl, der nur bei der 30sten und zwischen der 50sten und 70sten Generation abfällt. Bei der 60sten Generation besteht ein Großteil aller Individuen aus nur drei Unterverteilungen. Über die

Generationen wandelt sich dies wieder bis zur 70sten Generation, bei der dann die Mehrzahl der Individuen aus sieben Unterverteilungen besteht. Dies kann einerseits durch die zufällige Paarungsselektion ausgelöst werden, bei der Individuen mit sieben Unterverteilungen häufiger ausgewählt werden. Andererseits können sich geringe Verbesserungen der Fitness bei der Umweltselektion niederschlagen, bei der dann Individuen mit einer anderen Unterverteilungsanzahl bevorzugt selektiert werden.

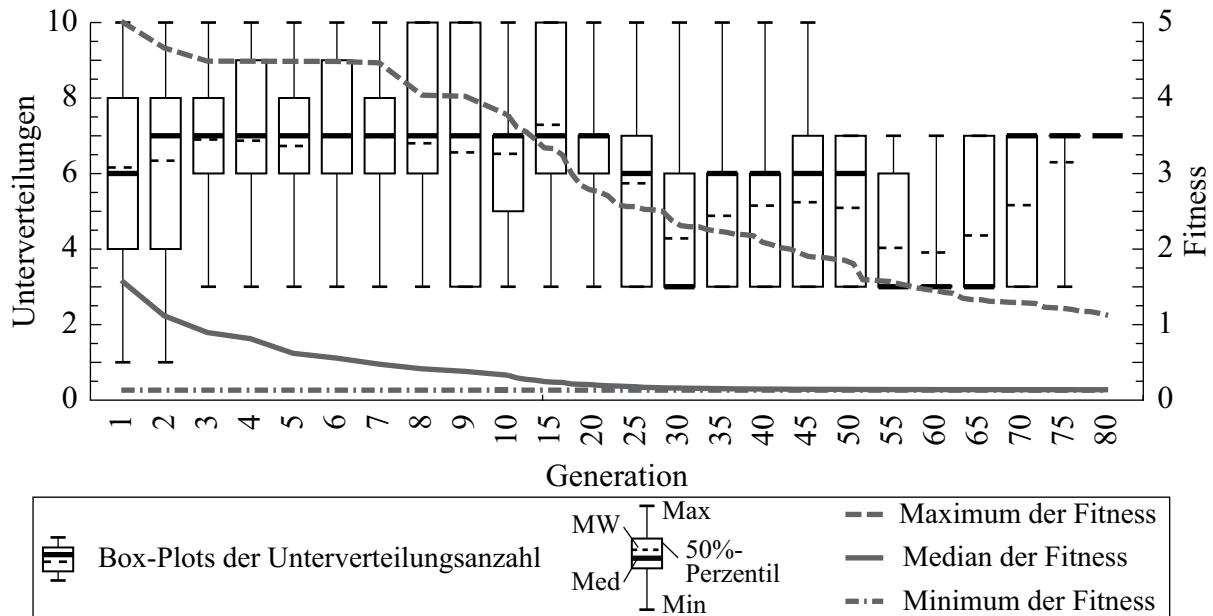


Bild 7.5: Verlauf der Häufigkeiten der Unterverteilungen und der Fitness über den Generationen bei der Schätzung des Seriensystems Typ II_s

Am Verlauf des Maximums der Fitness ist eine kontinuierliche Verbesserung bis zur 80sten Generation zu erkennen. Im nicht dargestellten Bereich bis zur 100sten Generation setzt sich dies fort. Beim Verlauf des Medians der Fitness ist spätestens ab der 40sten Generation nahezu keine Verbesserung mehr zu erkennen, was darauf hindeutet, dass sich die Individuen soweit aneinander angenähert haben, dass kaum mehr eine Verbesserung zu erwarten ist. Die Güte der Schätzung ist damit von der Anzahl der Unterverteilungen spätestens ab diesem Zeitpunkt nahezu unabhängig. Welche Unterverteilungsanzahl zwischen drei und zehn am Ende der Replikation auftritt, ist damit beinahe zufällig.

7.1.3 Möglichkeiten zur Verbesserung der Schätzung der Unterverteilungsanzahl

Durch den zusätzlichen freien Parameter bei der Schätzung der Unterverteilungsanzahl steigt die Vielfalt an Lösungsmöglichkeiten extrem an. Dies führt aber nicht zwangsläufig zu einer verbesserten Schätzung. Vielmehr verringert sich die Anpassung an die Datenpunkte, da nur Individuen mit gleicher Anzahl an Unterverteilungen bei der Rekombination berücksichtigt werden können. Dadurch sinkt die Anzahl an Individuen

für jede mögliche Unterverteilungsanzahl, wodurch eine schlechtere Abdeckung des Lösungsraumes je Unterverteilungsanzahl möglich ist.

Sowohl bei Mischverteilungen als auch bei Systemen tritt das Phänomen auf, dass überflüssige Unterverteilungen geschätzt werden. Dies führt zu einer Erhöhung der Unterverteilungsanzahl, aber nicht zu einer Verbesserung der Güte der Schätzung. Bei Mischverteilungen ist eine Möglichkeit, dieses Phänomen zu vermindern, die Eingrenzung des Lösungsintervalls des Gewichtungsfaktors p . Damit kann verhindert werden, dass Verteilungen mit sehr geringer Gewichtung in der Schätzung auftreten. In Bild 7.6 ist die Verteilung der Anzahl der Unterverteilungen für Mischverteilungstyp E dargestellt. Das Lösungsintervall des Gewichtungsfaktors p wurde auf $[0,05, 0,75]$ geändert.

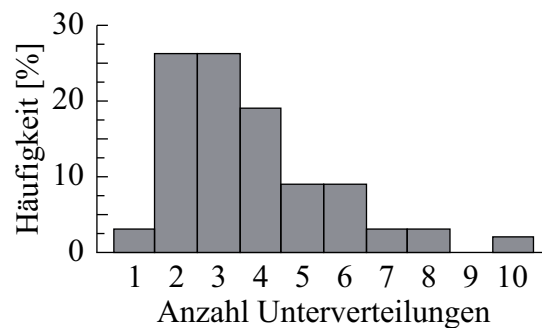


Bild 7.6: Verteilung der Anzahl der Unterverteilungen bei 100 Replikationen nach der letzten Generation bei der Schätzung der Mischverteilung Typ E mit angepasstem Intervall der Gewichtung

Gut zu erkennen ist, dass die Anzahl an Schätzungen mit einer Unterverteilung stark abnimmt. Dagegen nehmen die Schätzungen mit zwei und drei Unterverteilungen zu. Bei der Einschränkung des Intervalls der Gewichtung muss darauf geachtet werden, dass nicht alle Verteilungen mit geringen Werten von p ausgeschlossen werden.

Bei Seriensystemen besteht die Möglichkeit, durch eine Straf- oder Penaltyfunktion die Schätzungen mit großer Unterverteilungsanzahl durch eine höhere Fitness zu bestrafen. Dabei muss jedoch darauf geachtet werden, nicht zu früh die Vielfalt an möglichen Unterverteilungen einzuengen, um den Algorithmus nicht zu einer schlechten Lösung mit wenigen Unterverteilungen konvergieren zu lassen.

Sowohl bei Mischverteilungen als auch bei booleschen Systemen wäre es möglich, während der Berechnung die Unterverteilungen zu identifizieren, die keinen Einfluss auf die Mischverteilungs- oder Systemkurve besitzen. Werden diese direkt aus der Kodierung des Individuums entfernt, können bessere Schätzungen mit einer realistischen Anzahl an Unterverteilungen erwartet werden.

Ungünstig bei der Bestimmung der Unterverteilungsanzahl ist die eingeschränkte Möglichkeit zur Rekombination. Dadurch werden die Möglichkeiten zur Abdeckung des gesamten Lösungsraumes stark verringert. Um dieser Einschränkung zu begegnen, müsste eine Möglichkeit geschaffen werden, die die Rekombination aller Individuen erlauben würde. Dies kann erreicht werden, indem die fehlenden oder überzähligen Unterverteilungen eines Individuums gegenüber dem zuerst zur Rekombination selektierten sinnvoll ergänzt oder entfernt werden. Dazu müssen die einzelnen Unterverteilungen innerhalb eines Individuums bewertet werden. Bei dieser Bewertung muss die Ähnlichkeit der Unterverteilungen der Individuen eine zentrale Rolle spielen.

7.2 Bewertung der Anwendung

Anhand der in Kapitel 5 vorgestellten repräsentativen gemischten Verteilungen konnte gezeigt werden, dass der genetische Algorithmus prinzipiell gut zur Parameterschätzung geeignet ist. Zusätzlich konnte anhand der betrachteten Fälle das Verhalten des Algorithmus bei verschiedenen Problemstellungen untersucht werden. Im Folgenden wird die Anwendbarkeit des genetischen Algorithmus zur Parameterschätzung von gemischten Verteilungen im Hinblick auf verschiedene Problematiken diskutiert.

Genetischer Algorithmus als stochastische Methode (Heuristik)

Der große Vorteil von stochastischen Methoden ist ihr einfacher Aufbau und daraus folgend die Variabilität zum Einsatz für verschiedene Problemstellungen. Bei allen stochastischen Methoden ist die Güte des Ergebnisses ein Produkt des Zufalls. Dadurch ist ein erzielttes Ergebnis nicht zwangsläufig ein Optimum. In den meisten Fällen ist die Güte der erzielbaren Ergebnisse jedoch ausreichend. Die Anwendung des genetischen Algorithmus auf die Schätzung von gemischten Verteilungen zeigt bei mehreren Replikationen in den meisten Fällen gute Ergebnisse. Ausreißer zum Positiven und Negativen sind bei komplexeren Problemen nahezu immer vorhanden. Die Durchführung eines einzelnen Durchlaufs führt jedoch nur bei einfachen Problemstellungen, wie bei der Schätzung von Einfachverteilungen, zu ausreichend guten Schätzungen.

Fitnessfunktion

Die Anwendung des genetischen Algorithmus gründet sich auf den Vergleich der Güte von Individuen anhand einer vordefinierten Fitnessfunktion. Diese Fitnessfunktion muss geeignet sein, die Individuen eindeutig entsprechend ihrer Güte zu beschreiben. Dazu ist die Vorkenntnis der Art der zu schätzenden Verteilung und vor allem die Kenntnis der Datenherkunft, aus einer Mischverteilung oder einem System, erforderlich. Eine fehlende Kenntnis führt zu einer schlechten Anpassung an die Daten und zu falschen Aussagen über das Ausfallverhalten. Ohne vorherige Analyse der Daten ist ein Einsatz der Methode nicht sinnvoll. Zur Analyse der Daten vor der Durchführung

der Parameterschätzung sollten die Punkte zum Auftreten von gemischten Verteilungen aus Kapitel 2.2.3 beachtet werden.

Parameter des Algorithmus

Die universelle Anwendbarkeit und Einsetzbarkeit der genetischen Algorithmen basiert auf einer Vielzahl von Parametern, mit denen der Algorithmus auf die jeweilige Problemstellung angepasst werden kann. Diese Anpassbarkeit führt dazu, dass ein allgemeingültiger Parametersatz zur Bearbeitung von definierten Problemstellungen gefunden werden kann. Dieser Parametersatz führt jedoch nicht immer zu optimalen Lösungen. Eine genaue Anpassung auf die jeweils vorliegende Problemstellung kann zu deutlich besseren Ergebnissen führen. Hierbei müssen vor allem die Abhängigkeiten der Parameter untereinander beachtet werden. Die zur Schätzung der Verteilungsparameter in den repräsentativen Datensätzen verwendeten Parameter des Algorithmus sind ein Kompromiss aus allgemeiner Anwendbarkeit und Güte der Ergebnisse.

Datensätze und Verteilungen

Parameterschätzungen mit genetischen Algorithmen sind stark von den Datensätzen und den zugrunde liegenden Verteilungen abhängig. Die Güte der Parameterschätzungen wird, bei ausreichender Stichprobengröße, wenig von stochastischen Einflüssen innerhalb des Datensatzes beeinflusst. Sehr viel stärker beeinflusst die Charakteristik der zugrundeliegenden Verteilungen das Ergebnis. Problematisch sind hierbei vor allem Verteilungen mit Parametern, die mehrere Zehnerpotenzen auseinanderliegen. Bei diesen ist die Schätzungsgüte stark von der gewählten Genauigkeit der Parameter und dem Lösungsraum abhängig.

Schätzung der Unterverteilungsanzahl

Die Schätzung der Unterverteilungsanzahl durch den genetischen Algorithmus bietet die Möglichkeit, in vorhandenen Datensätzen nicht nur die Parameter der Unterverteilungen, sondern auch deren Anzahl zu bestimmen. Dadurch ist es möglich, ohne direkte Vorgabe eine Parameterschätzung von gemischten Verteilungen durchzuführen. In der bisherigen Umsetzung sind Schätzungen der Unterverteilungsanzahl möglich, jedoch müssen die Ergebnisse einer kritischen Prüfung auf Plausibilität unterzogen werden. Dabei ist das Hauptaugenmerk auf die Schätzungen von überflüssigen Verteilungen außerhalb des für die Gesamtverteilung maßgeblichen Bereichs zu legen.

8 Zusammenfassung und Ausblick

Der Nachweis von Zuverlässigkeitsanforderungen erfordert einen großen Zeit- und Kostenaufwand. Einerseits kann dieser Nachweis durch Prüfstandsversuche, andererseits durch die Auswertung von Felddaten erreicht werden. Beide Möglichkeiten haben Vor- und Nachteile. Versuche bieten den Vorteil der genauen Kenntnis der Parameter auf Kosten meist kleiner Stichprobengrößen und synthetischer Versuchsbedingungen. Felddaten bieten den Vorteil der realen Belastung des Produkts und großer Stichprobenumfänge mit den Nachteilen der Unsicherheit bezüglich der genauen Kenntnis der Lebensdauer und der erfahrenen Belastung. Trotz der genannten Nachteile bietet die Analyse von Felddaten ein großes Potential wegen der Kenntnis von Lebensdauern unter realer Belastung und den realen Ausfallmechanismen der Produkte. Felddatensätze bestehen meist aus einer Mischung verschiedener Produkttypen oder Ausfallmechanismen. Diese Mischungen können durch Mischverteilungen oder Ausfälle verschiedener Komponenten von Systemen auftreten. Zur Analyse dieser gemischten Verteilungen wurde in dieser Arbeit ein neues Verfahren vorgestellt.

Dazu wurden zunächst die bekannten grafischen und analytischen Verfahren zur Analyse von Weibull-Mischverteilungen erläutert. Für die Problemstellung der Analyse von Mischverteilungen und booleschen Systemen wurden die Einschränkungen der vorgestellten Methoden aufgezeigt. Im Weiteren wurden bekannte numerische Optimierungsmethoden und deren Arbeitsweise und Anwendung beschrieben. Diese numerischen Optimierungsmethoden sind auch dann einsetzbar, wenn die Verwendung analytischer Methoden nicht möglich ist.

Aus den vorgestellten Optimierungsmethoden wurden zwei vielversprechende Verfahren ausgewählt, die genetischen Algorithmen und die künstlichen neuronalen Netze. Diese wurden genauer auf ihre Anwendbarkeit auf die Problemstellung untersucht. Dazu wurden zunächst der Ablauf und die Anwendung der genetischen Algorithmen als Optimierungsmethode beschrieben. Ebenfalls wurden künstliche neuronale Netze und ihr Aufbau sowie deren Anwendung vorgestellt. Die beiden Methoden wurden anschließend bezüglich der vorliegenden Problemstellung, der Parameterschätzung von gemischten Verteilungen, gegenübergestellt. Aus der Diskussion der beiden Methoden konnten die genetischen Algorithmen aufgrund ihrer universelleren Anwendbarkeit als die geeignetere Variante ausgewählt werden.

Im nächsten Schritt wurde ein genetischer Algorithmus zur Schätzung von gemischten Verteilungen implementiert. Die Analyse der verschiedenen Algorithmenparameter

und deren Optimierung konnte anhand von beispielhaften repräsentativen Mischverteilungen durchgeführt werden. Hierbei lag das Hauptaugenmerk auf den zentralen Punkten der Kodierung der Individuen, der geeigneten Ziel- oder Fitnessfunktion und den optimalen Parametern des genetischen Algorithmus. Aufbauend auf diesen Untersuchungen wurde eine Empfehlung für die Parameter zur Schätzung von Mischverteilungen gegeben.

Zur Untersuchung der Funktionsweise des Algorithmus und der zu erwartenden Ergebnisse wurden verschiedene repräsentative Datensätze erstellt. Die genetischen Algorithmen als stochastische Methode können ein globales Optimum jedes Optimierungsproblems finden, dazu müssen jedoch sehr große Populationen und viele Generationen durchlaufen werden. Anhand der vorgestellten Datensätze wurde untersucht, welche Ergebnisse bei einer begrenzten Populationsgröße und Generationenanzahl erwartet werden können. Dazu mussten die Datensätze möglichst viele Arten von Mischverteilungen und booleschen Systemen repräsentieren, um allgemeine Aussagen treffen zu können. Aufbauend auf den repräsentativen Datensätzen wurden Parameterschätzungen durchgeführt und deren Ergebnisse diskutiert. Bei Mischverteilungen ohne ausfallfreie Zeit wurden Ergebnisse erzielt, die im Bereich der Ergebnisse kommerzieller Software liegen. Mit zunehmender Anzahl an Unterverteilungen konnte die Ergebnisgüte der kommerziellen Software nicht mehr in allen Fällen erreicht werden. Bei der Schätzung von Mischverteilungen mit ausfallfreier Zeit und booleschen Systemen konnten gute bis sehr gute Ergebnisse erzielt werden. In einem weiteren Schritt wurde die Möglichkeit zur Schätzung der Unterverteilungsanzahl untersucht, wurden die Schwierigkeiten erläutert und verschiedene Verbesserungsmaßnahmen vorgeschlagen.

Mit dem genetischen Algorithmus konnte eine Methode implementiert werden, mit der es möglich ist, die Verteilungsparameter aller Typen von gemischten Verteilungen zu schätzen. Die Güte der Ergebnisse reicht an die der analytischen Vergleichsmethode MLE/EM heran. Bei booleschen Systemen war bisher eine Parameterschätzung der Unterverteilungen überhaupt nicht möglich. Die universelle Anwendbarkeit der Methode auf alle Arten von gemischten Verteilungen gleicht den Nachteil der etwas geringeren Ergebnisgüte gegenüber den analytischen Methoden aus.

In dieser Arbeit konnte die prinzipielle Eignung numerischer Methoden zur Parameterschätzung von gemischten Verteilungen gezeigt werden. Darauf aufbauend könnten weitere numerische Methoden auf ihre Anwendbarkeit auf diese Problemstellung hin untersucht werden. Der genetische Algorithmus zeigt für die vorgestellten Verteilungen gute bis sehr gute Ergebnisse. Ein Nachweis der Anwendbarkeit auf weitere Verteilungsarten wie beispielsweise die Lognormalverteilung steht noch aus. Zudem sind noch weitere Möglichkeiten zur Erweiterung des genetischen Algorithmus bekannt bis hin zu memetischen Algorithmen, die bisher nicht untersucht wurden.

9 Literatur

- [1] Bertsche, B.; Lechner, G.: Zuverlässigkeit im Fahrzeug- und Maschinenbau. Berlin: Springer, 2004
- [2] Delonga, M.: Zuverlässigkeitsmanagement auf Basis von Felddaten. Stuttgart, Universität, Institut für Maschinenelemente, Diss., 2006
- [3] Edler, A.: Nutzung von Felddaten in der qualitätsgetriebenen Produktentwicklung und im Service. Berlin, Technische Universität, Institut für Werkzeugmaschinen und Fabrikbetrieb, Diss., 2001
- [4] Bertsche, B.; Göhner, P.; Jensen, U.; Schinköthe, W.; Wunderlich, H.-J.: Zuverlässigkeit mechatronischer Systeme. Grundlagen und Bewertungen in frühen Entwicklungsphasen. Berlin: Springer, 2009
- [5] Kao, J. H. K.: A Graphical Estimation of Mixed Weibull Parameters in Life-Testing of Electron Tubes. In: Technometrics, Vol. 1, No. 4, 1959
- [6] Cran, G. W.: Graphical Estimation Methods for Weibull distributions. In: Microelectronics & Reliability, Vol. 15, 1976
- [7] Jensen, F.; Petersen, N. E.: Burn-In - An Engineering Approach to the Design and Analysis of Burn-in Procedures. Chichester: John Wiley & Sons, 1982
- [8] Weicker, K.: Evolutionäre Algorithmen. Wiesbaden: Teubner, 2007
- [9] Jiang, S.; Kececioglu, D.: Graphical representation of two mixed-Weibull distributions. In: IEEE Transactions on Reliability, Vol. 41, No. 2, 1992
- [10] O'Connor, P. D. T.: Zuverlässigkeitstechnik: Grundlagen und Anwendungen. Weinheim: VCH, 1990
- [11] Kececioglu, D.: Reliability & Life Testing Handbook, Volume 1. London: PTR Prentice Hall, 1993
- [12] Kececioglu, D.: Reliability & Life Testing Handbook, Volume 2. London: PTR Prentice Hall, 1994
- [13] Birolini, A.: Reliability Engineering: Theory and Practice. Berlin: Springer, 2004
- [14] Kozlov, B. A.; Ušakov, I. A.: Handbuch zur Berechnung der Zuverlässigkeit für Ingenieure. München: Hanser, 1979

- [15] Verband der Automobilindustrie e.V. (Hrsg.): Qualitätsmanagement in der Automobilindustrie - Zuverlässigkeitssicherung bei Automobilherstellern und Lieferanten. Band 3.2, 3. überarbeitete und erweiterte Auflage. Frankfurt: VDA, 2004
- [16] Abernathy, R. B.: The new Weibull Handbook. Selbstverlag, 2000
- [17] Härtler, G.: Statistische Methoden für die Zuverlässigkeitsanalyse. Berlin: Springer, 1981
- [18] Dhillon, B. S.: Mechanical Reliability: Theory, Models and Applications. Washington, D.C.: American institute of aeronautics and astronautics inc., 1988
- [19] Beljaev, J. K.; Kahle, W.: Methoden der Wahrscheinlichkeitsrechnung und Statistik bei der Analyse von Zuverlässigkeitsdaten. Stuttgart: Teubner, 2000
- [20] Ebeling, C. E.: An introduction to Reliability and Maintainability Engineering. New York: McGraw Hill, 1997
- [21] Meeker, Q. W.; Escobar, L. A.: Statistical Methods for Reliability Data. New York: John Wiley & Sons, 1998
- [22] Lawless, J. F.: Statistical Models and Methods for lifetime Data. New York: John Wiley & Sons, 1982
- [23] Verein Deutscher Ingenieure e. V.: VDI Richtlinie 4001 Blatt 2: Terminologie der Zuverlässigkeit. Düsseldorf: VDI, 2006
- [24] Norm DIN 40041: 1990. Zuverlässigkeit – Begriffe
- [25] Wilker, H.: Weibull-Statistik in der Praxis – Band 3: Leitfaden zur Zuverlässigkeitsermittlung technischer Produkte. Norderstedt : Books on Demand, 2004
- [26] Benard, A.; Bos-Levenbach, E. C.: Het uitzetten van waarnemingen op waarschijnlijkheids-papier. In: Statistica Neerlandica, Vol. 7, No. 3, 1953
- [27] Meyna, A.; Pauli, B.: Zuverlässigkeitsprognosen für Kfz-Komponenten bei unvollständigen Daten. In: Automobiltechnische Zeitschrift, Heft 12, 2000
- [28] Johnson, L. G.: The Statistical Treatment of Fatigue Experiments. New York: Elsevier publishing Company, 1964
- [29] Eckel, G.: Bestimmung des Anfangsverlaufs der Zuverlässigkeitsfunktion von Automobilteilen. In: Qualität und Zuverlässigkeit, Jahrg. 22, Heft 9, 1977
- [30] Kalbfleisch, J. D.; Prentice, R. L.: The statistical Analysis of Failure Time Data. New York: John Wiley & Sons, 1980

- [31] Gaede, K.-W.: Zuverlässigkeit Mathematische Modelle. München: Hanser, 1977
- [32] Maillart, L. M.: Introduction to Markov-Chain Modeling, Analysis and Optimization. Tutorial Notes RAMS 2010, San Jose, USA, 2010
- [33] Nelson, W.: Applied Life Data Analysis. New York: John Wiley & Sons, 1982
- [34] Remy, E.; Idée, E.; Briand, P.; François, R.: Bibliographical review and numerical comparison of statistical estimation methods for the three-parameter Weibull distribution. Proc. ESREL 2010, Rhodos, Griechenland, 2010
- [35] Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Berlin: Springer, 2003
- [36] Hochstättler, W.: Algorithmische Mathematik. Berlin: Springer, 2010
- [37] Mundry, E.: On the Resolution of Mixed Frequency Distributions into Normal Components. In: Mathematical Geology, Vol. 4, No. 1, 1972
- [38] Nothnagel, K.: Ein graphisches Verfahren zur Zerlegung von Mischverteilungen. In: Qualitätskontrolle, Jahrg. 13, Heft 2, 1968
- [39] Harding, J. P.: The use of probability paper for the graphical analysis of polymodal frequency distributions. In: Journal of the Marine Biological Association of the United Kingdom, Vol. 28, No. 1, 1949
- [40] Cassie, R. M.: Some uses of Probability paper in the Analysis of size Frequency Distributions. In: Australian Journal of Marine and Freshwater Research, Vol. 5, No. 3, 1954
- [41] Jiang, S.: Mixed Weibull distributions in reliability engineering - statistical models for the lifetime of units with multiple modes of failure. Tucson, University of Arizona, Aerospace and Mechanical Engineering Graduate College, Diss., 1991
- [42] Falls, L. W.: Estimation of Parameters in Compound Weibull Distributions. In: Technometrics, Vol. 12, No. 2, 1970
- [43] Cheng, S. W.; Fu, J. C.: Estimation of mixed Weibull parameters in life testing. In: IEEE Transactions on Reliability, Vol. 31, No. 4, 1982
- [44] Böhning, D.: Computer-Assisted Analysis of Mixtures and Applications - Meta-analysis, disease mapping and others. Boca Raton: Chapman & Hall/CRC, 2000

- [45] Lindsay, B. G.: Mixture Models: Theory, Geometry and Applications. In: NSF-CBMS Regional Conference Series in Probability and Statistics , Vol. 5, 1995
- [46] Everitt, B.: Introduction to Optimization Methods and Application in Statistics. London: Chapman and Hall, 1987
- [47] Moré, J. J.: The Levenberg-Marquardt Algorithm: Implementation and Theory. In: Lecture Notes in Mathematics, Numerical Analysis Vol. 630, 1978
- [48] Weise, T.: Global Optimization Algorithms - Theory and Application. Online: <http://www.it-weise.de/projects/book.pdf>, Stand 03.02.2012
- [49] Powell, M. J. D.: A new algorithm for unconstrained optimization. Nonlinear Programming. New York: Academic Press, 1970
- [50] Moré, J. J.: Recent developments in algorithms and software for trust region methods. Mathematical Programming: The State of the Art. Berlin: Springer, 1983
- [51] Bracke, S.: Contribution for a sustainable design of complex mechatronic automotive products using statistical damage prediction models in the early product construction phase to prevent damage causes. Proc. Global Conference on Sustainable Product Development and Life Cycle Engineering 2008, Pusan, Korea, 2008
- [52] Geiger, C.; Kanzow, C.: Theorie und Numerik restringierter Optimierungsaufgaben. Berlin: Springer, 2002
- [53] Levitin, G.; Hu, X.; Dai, Y.-S.: Particle Swarm Optimization in Reliability Engineering. In: Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence, Vol. 40, 2007
- [54] Liang, Y.-C.; Smith, A. E.: The Ant Colony Paradigm for Reliable Systems Design. In: Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence, Vol. 40, 2007
- [55] Agarwal, M.; Sharma, V. K.: Ant Colony Optimization Algorithm for Heterogeneous Redundancy Allocation in Multi-State Series-Parallel Systems. In: International Journal of Reliability, Quality and Safety Engineering, Vol. 16, No. 2, 2009
- [56] Sharman, K. C.: Maximum Likelihood Parameter Estimation by Simulated Annealing. Proc. ICASSP 1988, New York, USA, 1988
- [57] Sharman, K. C.; McClurkin, G. D.: Genetic Algorithms for Maximum Likelihood Parameter Estimation. Proc. ICASSP 1989, Glasgow, UK, 1989

- [58] Czepiel, S. A.: Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. Online: <http://www.czep.net/stat/mlelr.pdf>, Stand 02.02.2012
- [59] Fritz, A.: Berechnung und Monte-Carlo Simulation der Zuverlässigkeit und Verfügbarkeit technischer Systeme. Stuttgart, Universität, Institut für Maschinenelemente, Diss., 2001
- [60] Rao, C. R.: Was ist Zufall? – Statistik und Wahrheit: München: PTR Prentice Hall, 1995
- [61] Pozsgai, P.: Realitätsnahe Modellierung und Analyse der operativen Zuverlässigkeitskennwerte technischer Systeme. Stuttgart, Universität, Institut für Maschinenelemente, Diss., 2004
- [62] Darwin, C. R.: The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (6th ed.). London: John Murray, 1872
- [63] Eiben, A. E.; Smith, J. E.: Introduction to Evolutionary Computing. Berlin: Springer, 2003
- [64] Holland, J. H.: Genetic Algorithms and the optimal allocation of trials. In: SIAM Journal of Computing, Vol. 2, 1973
- [65] Rechenberg, I.: Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der Biologischen Evolution. Stuttgart: Frommann-Holzboog, 1973
- [66] Koza, J. R.: Genetic Programming. On the Programming of Computers by Means of Natural Selection. Cambridge: The MIT Press, 1992
- [67] Fogel, L. J.; Owens, A. J.; Walsh, M. J.: Artificial intelligence through simulated evolution. New York: John Wiley & Sons, 1966
- [68] Hart, W. E.; Krasnogor, N.; Smith, J. E.: Recent Advances in Memetic Algorithms. Berlin: Springer, 2005
- [69] Merz, P.: The Compact Memetic Algorithm. Proc. GECCO 2003, Chicago, USA, 2003
- [70] Hart, W. E.; Krasnogor, N.; Smith, J. E.: Memetic Evolutionary Algorithms. In: Studies in Fuzziness and Soft Computing - Recent Advances in Memetic Algorithms, Vol. 166. Berlin: Springer, 2005
- [71] Marseguerra, M.; Zio, E.; Martorell, S.: Basics of genetic Algorithms optimization for RAMS applications. In: Reliability Engineering & System Safety, Vol. 91, No. 9, 2006

- [72] Martorell, S.; Carlos, S.; Sánchez, A.; Serradell, V.: Constrained optimization of test intervals using a steady-state genetic algorithm. In: *Reliability Engineering & System Safety*, Vol. 67, No. 3, 2000
- [73] Martorell, S.; Carlos, S.; Villanueva, J. F.; Sánchez, A.: Genetic Algorithm Applications in Surveillance and Maintenance Optimization. In: *Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence*, Vol. 39, 2007
- [74] Thomas, G. M.; Gerth, R.; Velasco, T.; Rabelo, L. C.: Using real-coded genetic Algorithms for Weibull Parameter Estimation. In: *Computers & Industrial Engineering*, Vol. 29, No. 1 - 4, 1995
- [75] Jenab, K.; Kazeminia, A.; Liu, D. S.: A computational intelligent approach to estimate the Weibull parameters. In: *International Journal of Industrial and Systems Engineering*, Vol. 6, No. 1, 2010
- [76] Tolson, B. A.; Maier, H. R.; Simpson, A. R.; Lence, B. J.: Genetic Algorithms for Reliability-Based Optimization of Water Distribution Systems. In: *Journal of Water Resources Planning and Management*, Vol. 130, No. 1, 2004
- [77] Shayanfar, M. A.; Massah, S. R.; Rahami, H.: An Inverse Reliability Method Using Neural Networks and Genetic Algorithms. In: *World Applied Sciences Journal*, Vol. 2 No. 6, 2007
- [78] Huang, H.-Z.; Tian, Z.; Zuo, M. J.: Optimal Redundancy Allocation of Multi-State Systems with Genetic Algorithms. In: *Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence*, Vol. 39, 2007
- [79] Marseguerra, M.; Zio, E.; Podofillini, L.: Genetic Algorithms and Monte Carlo Simulation for the Optimization of System Design and Operation. In: *Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence*, Vol. 39, 2007
- [80] Davis, D. D.; De Jong, K.; Vose, M. D.; Whitley, L. D. (Hrsg.): *Evolutionary Algorithms*. New York: Springer, 1999
- [81] Quagliarella, D.; Périaux, J.; Poloni, C.; Winter, G. (Hrsg.): *Genetic Algorithms and Evolution Strategies in Engineering and Computer Science: Recent Advances and Industrial Applications*. Chichester: John Wiley & Sons, 1998
- [82] Goldberg, D. E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, Mass.: Addison-Wesley, 1998

- [83] Hamming, R. W.: Error Detecting and Error Correcting Codes. In: Bell System Technical Journal, Vol. 29, 1950
- [84] Nissen, V.: Einführung in Evolutionäre Algorithmen. Braunschweig: Vieweg, 1997
- [85] Bäck, T.: Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. New York: Oxford University Press, 1996
- [86] Heistermann, J.: Genetische Algorithmen. Theorie und Praxis evolutionärer Optimierung. Stuttgart: Teubner, 1994
- [87] Mühlenbein, H.: How genetic algorithms really work: Mutation and hill-climbing. Proc. PPSN 1992, Brüssel, Belgien, 1992
- [88] Jansen, T.: Theoretische Analyse evolutionärer Algorithmen unter dem Aspekt der Optimierung in diskreten Suchräumen. Dortmund, Universität, Fachbereich Informatik, Diss., 2000
- [89] Kosiński, W.; Kotowski, S.; Socała, J.: On Asymptotic Behaviour of a Binary Genetic Algorithm. Proc. XXI Autumn Meeting of Polish Information Processing Society 2005, Wisła, Polen, 2005
- [90] Kosiński, W.; Kotowski, S.; Socała, J.: On Asymptotic Behaviour of a Simple Genetic Algorithm. In: Intelligent Information Processing and Web Mining, Advances in Soft Computing, Vol. 35, 2006
- [91] Schaefer, R.: Foundations of Global Genetic Optimization. In: Studies in Computational Intelligence, Vol. 74, 2007
- [92] Rudolph, G.: Convergence Analysis of Canonical Genetic Algorithms. In: IEEE Transactions on Neural Networks, Vol. 5, No. 1, 1994
- [93] Eiben, A. E.; Aarts, E. H. L.; van Hee, K. M.: Global Convergence of Genetic Algorithms: a Markov Chain Analysis. In: Parallel Problem Solving from Nature, Lecture Notes in Computer Science 496. Berlin: Springer, 1991
- [94] McCulloch, W. S.; Pitts, W.: A logical Calculus of the ideas immanent in nervous Activity. In: Bulletin of Mathematical Biophysics, Vol. 5, 1943
- [95] Hebb, D. O.: The organisation of behavior. New York: John Wiley & Sons, 1949
- [96] Widrow, B.; Hoff, M. E.: Adaptive Switching Circuits. In: IRE WESCON Convention Record, Part 4, 1960
- [97] Rumelhart, D. E.; Hinton, G. E.; Williams, R. J.: Learning internal representations by error propagation. Parallel Distributed Processing: Explorations in

- the Microstructure of Cognition. Volume 1: Foundations, Cambridge: MIT Press, 1986
- [98] Scherer, A.: Neuronale Netze - Grundlagen und Anwendungen. Braunschweig: Vieweg, 1997
- [99] Ritter, H.; Martinetz, T.; Schulten, K.: Neuronale Netze - Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke. Bonn: Addison-Wesley, 1991
- [100] Schöneburg, E.; Hansen, N.; Gawelczyk, A.: Neuronale Netzwerke - Einführung, Überblick und Anwendungsmöglichkeiten. Haar bei München: Markt- u. Technik-Verl., 1990
- [101] VDI e. V.; VDE e. V. (Hrsg.): VDI/VDE Richtlinie 3550 Blatt 1: Computational Intelligence - Künstliche Neuronale Netze in der Automatisierungstechnik - Begriffe und Definitionen. Berlin: Beuth, 2001
- [102] Rojas, R.: Neural Networks - A Systematic Introduction. Berlin: Springer, 1996
- [103] Van Mahnen, J.: Beitrag zur neuronalen Schätzung von Verteilungsparametern in der Technischen Zuverlässigkeit. VDI-Fortschritt-Bericht, Reihe 20: Rechnergestützte Verfahren, Nr.269, Düsseldorf: VDI-Verlag, 1998
- [104] Fink, O.; Weidmann, U.; Hofmann, D.; Krolo, A.: Anwendungspotential künstlicher neuronaler Netze in der Prognose und Analyse der technischen Zuverlässigkeit. Entwicklung und Betrieb zuverlässiger Produkte, VDI-Berichte 2146, TTZ 2011, Stuttgart, 2011
- [105] Maisch, C.; Bertsche, B.: Prognose der Öllebensdauer mittels künstlicher Intelligenz. Entwicklung und Betrieb zuverlässiger Produkte, VDI-Berichte 2065, TTZ 2009, Stuttgart, 2009
- [106] Hu, Q. P.; Xie, M.; Ng, S. H.: Software Reliability Predictions using Artificial Neural Networks. In: Computational Intelligence in Reliability Engineering, Studies in Computational Intelligence, Vol. 40, 2007
- [107] Karunanithi, N.; Whitley, D.; Malaiya, Y. K.: Using neural networks in Reliability Prediction. In: Software, IEEE Vol. 9, No. 4, 1992
- [108] Coit, D. W.; Smith, A. E.: Solving the Redundancy allocation Problem using a combined Neural Network/ Genetic Algorithm approach. In: Computers and Operations Research - Special issue: artificial intelligence, evolutionary programming and operations research, Vol. 23 ,No. 6, 1996

- [109] Altiparmak, F.; Dengiz, B.; Smith, A. E.: A General Neural Network Model for Estimating Telecommunications Network Reliability. In: IEEE Transactions on Reliability, Vol. 58, No. 1, 2009
- [110] Sinha, S. K.; Pandey, M. D.: Probabilistic Neural Network for Reliability Assessment of Oil and Gas Pipelines. In: Computer-Aided Civil and Infrastructure Engineering, Vol. 17, No. 5, 2002
- [111] Liu, M. C.; Kuo, W.; Sastri, W.: An exploratory study of a neural network approach for reliability data analysis. In: Quality and Reliability Engineering International, Vol. 11, No. 2, 1995
- [112] Anders, U.: Statistische neuronale Netze. München: Verlag Franz Vahlen, 1997
- [113] Fahlmann, S. E.: An Empirical Study of Learning Speed in Back-Propagation Networks. Proc. Connectionist Models Summer School 1988, San Francisco, 1988
- [114] Riedmiller, M.; Braun, H.: Rprop - A Fast Adaptive Learning Algorithm. Proc. ISCIS VII 1992, Antalya, Türkei, 1992
- [115] Riedmiller, M.; Braun, H.: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. Proc. IEEE International Conference on Neural Networks 1993, Vol.1, San Francisco, USA, 1993
- [116] Klein, B.; Bertsche, B.: Estimating Parameters of Mixed Weibull Distributions using Genetic Algorithms. Proc. ESREL 2010, Rhodos, Griechenland, 2010
- [117] Klein, B.; Wolf, D.; Bertsche, B.: Bestimmung von Verteilungsparametern von Seriensystemen mit neuronalen Netzen. Entwicklung und Betrieb zuverlässiger Produkte, VDI-Berichte 2146, TTZ 2011, Stuttgart, 2011
- [118] Dietrich, E.; Schulze, A.: Statistische Verfahren zur Maschinen- und Prozessqualifikation. München: Hanser, 2009
- [119] Woodward, W. A.; Parr, W. C.; Schucany, W. R.; Lindsey, H.: A Comparison of Minimum Distance and Maximum Likelihood Estimation of a Mixture Proportion. In: Journal of the American Statistical Association, Vol. 79, No. 387, 1984
- [120] Bronstein, I. N.; Smendjajew, K. A.; Musiol, G.; Mühlig, H.: Taschenbuch der Mathematik. Frankfurt a. M.: Deutsch, 2001
- [121] Dubey, S. D.: Asymptotic Properties of Several Estimators of Weibull Parameters. In: Technometrics, Vol. 7, No. 3, 1965

- [122] Haibach, E.: Betriebsfestigkeit: Verfahren und Daten zur Bauteilberechnung. Berlin: Springer, 2006
- [123] Sachs, L.: Angewandte Statistik: Anwendung stochastischer Methoden. 9. überarb. Aufl. Berlin: Springer, 1999
- [124] Voß, W.: Taschenbuch der Statistik. Leipzig: Fachbuchverlag, München: Hanser, 2000
- [125] Koch, K. R.: Parameterschätzung und Hypothesentests in linearen Modellen. Bonn: Dümmler, 1997
- [126] Woodruff, B. W.; Moore, A. H.; Dunne, E. J.; Cortes, R.: A Modified Kolmogorov-Smirnov Test for Weibull Distributions with Unknown Location and Scale Parameters. In: IEEE Transactions on Reliability, Vol. 32, No. 2, 1983
- [127] Dufner, J.; Jensen, U.; Schumacher, E.: Statistik mit SAS. Stuttgart: Teubner, 2002
- [128] Agha, M.; Branker, D. S.: Maximum Likelihood Estimation and Goodness-of-fit Tests for Mixtures of Distributions. In: Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 46, No. 3, 1997
- [129] Claeskens, G.; Hart, J. D.: Goodness-of-fit tests in mixed models. In: Test, Vol. 18, No. 2, 2009
- [130] Molina, I.: Comments on: Goodness-of-fit tests in mixed modes. In: Test, Vol. 18, No. 2, 2009
- [131] Munk, A.; Krivobokova, T.: Comments on: Goodness-of-fit tests in mixed modes. In: Test, Vol. 18, No. 2, 2009
- [132] Claeskens, G.; Hart, J. D.: Rejoinder on: Goodness-of-fit tests in mixed models. In: Test, Vol. 18, No. 2, 2009
- [133] N. N.: Weibull++_The_mixed_weibull_distribution. Online: http://www.weibull.com/LifeDataWeb/the_mixed_weibull_distribution.htm, Stand 03.02.2012
- [134] Nocedal, J.; Wright, S. J.: Numerical Optimization. New York: Springer, 1999
- [135] Brodbeck, P.: Experimentelle und theoretische Untersuchungen zur Bauteilzuverlässigkeit und zur Systemberechnung nach dem Booleschen Modell. Stuttgart, Universität, Institut für Maschinenelemente, Diss., 1995

10 Anhang

10.1 Verwendete Parameter des genetischen Algorithmus

Tabelle A1.1: Initialisierungsparameter des Algorithmus und der Problemstellung für die Fitnessfunktion bei einer einzelnen Weibullverteilung

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Populationsgröße (E, N)	100,100	Eingangsdaten / Anzahl	Mediandaten / 50
Generationenanzahl	100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 200 \\ 0 & 100 \\ 1 & 1 \end{pmatrix}$
Kodierung	binär		
Rekombination	1-Punkt-Crossover		
Mutationswahrscheinlichkeit	0,001		
Selektion	Stochastisch	Genauigkeit der Parameter	(3, 2, 2, 5)
	Eltern + Nachkommen	Anzahl Unterverteilungen	(1, 1)
	Fitnessselektion		

Tabelle A1.2: Initialisierungsparameter des Algorithmus und der Problemstellung für die Fitnessfunktion bei einer Weibull-Mischverteilung

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Populationsgröße (E, N)	100,100	Eingangsdaten / Anzahl	Mediandaten / 100
Generationenanzahl	100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Kodierung	binär		
Rekombination	1-Punkt-Crossover		
Mutationswahrscheinlichkeit	0,001		
Selektion	Stochastisch	Genauigkeit der Parameter	(3, 5, 0, 5)
	Eltern + Nachkommen	Anzahl Unterverteilungen	(2, 2)
	Fitnessselektion		

Tabelle A1.3: Initialisierungsparameter des Algorithmus und der Problemstellung zum Vergleich der Rekombinationsvarianten

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	Mediandaten / 100
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	<i>variabel</i>		
Mutationswahrscheinlichkeit	0,001	Genauigkeit der Parameter	(3, 5, 0, 5)
Selektion	Stochastisch	Anzahl Unterverteilungen	(2, 2)
	Eltern + Nachkommen		
	Fitnessselektion		

Tabelle A1.4: Initialisierungsparameter des Algorithmus und der Problemstellung zum Vergleich der Mutationsvarianten

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	Mediandaten / 100
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	<i>variabel</i>	Genauigkeit der Parameter	(3, 5, 0, 5)
Selektion	Stochastisch	Anzahl Unterverteilungen	(2, 2)
	Eltern + Nachkommen		
	Fitnessselektion		

Tabelle A1.5: Initialisierungsparameter des Algorithmus und der Problemstellung zum Vergleich der Selektionsvarianten

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	Mediandaten / 100
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 0,018$	Genauigkeit der Parameter	(3, 5, 0, 5)
Selektion	<i>variabel</i>	Anzahl Unterverteilungen	(2, 2)
	<i>variabel</i>		
	<i>variabel</i>		

Tabelle A1.6: Initialisierungsparameter des Algorithmus und der Problemstellung zur Bestimmung des Einflusses der Generationenzahl, der Populationsgröße und der Lösungsraumgröße

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	Mediandaten / 100
Populationsgröße (E, N)	<i>variabel</i>	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} variabel \\ variabel \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	<i>variabel</i>		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 0,018$	Genauigkeit der Parameter	(var., var., 0, var.)
Selektion	stochastisch	Anzahl Unterverteilungen	(2, 2)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.7: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der Weibull-Einfachverteilungen ohne t_0

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / var
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 3 \\ 50 & 200 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(2, 2, 0, 0)
Selektion	Stochastisch	Anzahl Unterverteilungen	(1, 1)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.8: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der Weibull-Einfachverteilungen mit t_0

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / var
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 3 \\ 50 & 200 \\ 0 & 100 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(2, 2, 2, 0)
Selektion	Stochastisch	Anzahl Unterverteilungen	(1, 1)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.9: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der Mischverteilungen nach Jiang und Kececioğlu

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / var
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(3, 5, 0, 5)
Selektion	Stochastisch	Anzahl Unterverteilungen	(2, 2)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.10: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der Mischverteilungen mit mehreren Unterverteilungen

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	var / 100
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	150		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(3, 5, 0, 5)
Selektion	Stochastisch	Anzahl Unterverteilungen	(var, var)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.11: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der Mischverteilungen erweitert um t_0

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / 50
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 100 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	150		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(3, 5, 5, 5)
Selektion	Stochastisch	Anzahl Unterverteilungen	(2, 2)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.12: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der booleschen Serien- und Parallelsysteme ohne t_0

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / 50
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \end{pmatrix}$	$\begin{pmatrix} 0,5 & 10 \\ 0 & 1000 \\ 0 & 0 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover		
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Genauigkeit der Parameter	(3, 2, 0)
Selektion	Stochastisch	Anzahl Unterverteilungen	(2, 2)
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.13: Initialisierungsparameter des Algorithmus und der Problemstellung zur Parameterschätzung der booleschen Serien- und Parallelsysteme mit t_0

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / 50
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \end{pmatrix}$	$\begin{pmatrix} 0,5 & 10 \\ 0 & 1000 \\ 0 & 100 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover	Genauigkeit der Parameter	(3, 2, 2)
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Anzahl Unterverteilungen	(2, 2)
Selektion	Stochastisch		
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.14: Initialisierungsparameter des Algorithmus und der Problemstellung zur Schätzung der Unterverteilungen der Mischverteilungen

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / 100
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \\ p \end{pmatrix}$	$\begin{pmatrix} 0,1 & 5 \\ 0 & 500 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$
Generationenanzahl	150		
Kodierung	binär		
Rekombination	N-Punkt-Crossover	Genauigkeit der Parameter	(3, 5, 0, 5)
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Anzahl Unterverteilungen	(10, 10)
Selektion	Stochastisch		
	20 % Eltern+Nachkommen		
	Fitnessselektion		

Tabelle A1.15: Initialisierungsparameter des Algorithmus und der Problemstellung zur Schätzung der Unterverteilungen der booleschen Seriensysteme

Parameter Algorithmus	Einstellung	Parameter Problemstellung	Einstellung
Fitnessfunktion	Abszisse, Weibullgew.	Eingangsdaten / Anzahl	M-C Methode / 50
Populationsgröße (E, N)	100,100	Lösungsraumgröße $\begin{pmatrix} b \\ T \\ t_0 \end{pmatrix}$	$\begin{pmatrix} 0,5 & 10 \\ 0 & 1000 \\ 0 & 0 \end{pmatrix}$
Generationenanzahl	100		
Kodierung	binär		
Rekombination	N-Punkt-Crossover	Genauigkeit der Parameter	(3, 2, 0)
Mutationswahrscheinlichkeit	nach Jansen $p_{mut} = 1/l$	Anzahl Unterverteilungen	(10, 10)
Selektion	Stochastisch		
	20 % Eltern+Nachkommen		
	Fitnessselektion		

10.2 Mittels Monte-Carlo Methode erzeugte Datensätze

Tabelle A2.1: Datensatz mit $n = 20$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	5,09	5	20,32	9	43,22	13	67,19	17	191,75
2	7,90	6	27,43	10	46,08	14	67,50	18	230,31
3	11,79	7	27,44	11	51,38	15	72,00	19	233,01
4	13,16	8	41,21	12	54,00	16	97,30	20	289,66

Tabelle A2.2: Datensatz mit $n = 50$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,56	11	19,62	21	54,08	31	98,88	41	162,64
2	5,41	12	22,08	22	54,96	32	104,69	42	174,48
3	6,24	13	31,35	23	56,50	33	106,96	43	178,01
4	8,72	14	34,53	24	56,63	34	107,76	44	252,99
5	11,23	15	35,88	25	60,48	35	109,76	45	286,01
6	13,70	16	41,47	26	67,27	36	109,82	46	302,89
7	14,29	17	46,59	27	73,10	37	119,77	47	361,10
8	18,29	18	49,58	28	74,84	38	120,76	48	407,76
9	18,77	19	50,96	29	77,46	39	133,89	49	413,92
10	19,04	20	53,92	30	79,38	40	161,11	50	699,16

Tabelle A2.3: Datensatz mit $n = 100$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,79	21	17,86	41	38,20	61	77,85	81	122,66
2	2,47	22	18,18	42	40,84	62	81,13	82	128,35
3	3,72	23	18,42	43	43,20	63	81,60	83	134,96
4	6,12	24	18,69	44	44,57	64	81,72	84	135,69
5	7,69	25	19,24	45	48,02	65	91,00	85	136,14
6	8,05	26	20,39	46	50,55	66	94,12	86	141,93
7	9,69	27	21,85	47	51,45	67	96,94	87	142,43
8	10,37	28	23,83	48	51,48	68	97,42	88	144,35
9	10,38	29	23,85	49	54,93	69	99,90	89	148,16
10	11,20	30	24,53	50	55,01	70	106,04	90	173,49
11	11,52	31	27,51	51	55,22	71	106,59	91	178,63
12	13,34	32	28,91	52	56,10	72	109,66	92	204,88
13	13,39	33	29,81	53	56,14	73	111,87	93	223,42
14	13,52	34	34,28	54	60,10	74	113,86	94	224,80
15	13,94	35	34,84	55	60,26	75	114,43	95	245,11
16	14,42	36	35,56	56	61,23	76	115,23	96	274,47
17	15,32	37	37,60	57	62,22	77	115,69	97	306,74
18	16,68	38	37,95	58	72,68	78	119,22	98	314,51
19	16,77	39	38,04	59	75,64	79	121,30	99	406,49
20	17,57	40	38,20	60	75,76	80	121,48	100	441,74

Tabelle A2.4: Datensatz mit $n = 20$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,5B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	6,69	5	36,42	9	66,59	13	103,74	17	235,38
2	11,10	6	50,49	10	79,45	14	106,92	18	309,27
3	16,27	7	59,82	11	84,65	15	170,11	19	333,66
4	21,41	8	65,69	12	94,17	16	214,05	20	404,68

Tabelle A2.5: Datensatz mit $n = 50$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,5B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	5,85	11	33,94	21	53,22	31	79,51	41	140,39
2	8,41	12	35,06	22	54,03	32	84,12	42	171,47
3	10,66	13	36,05	23	58,62	33	87,13	43	191,65
4	14,71	14	37,83	24	59,68	34	91,37	44	192,65
5	18,68	15	37,84	25	60,66	35	102,07	45	209,52
6	22,83	16	39,64	26	62,07	36	103,21	46	219,31
7	24,49	17	40,16	27	62,39	37	105,32	47	239,32
8	31,50	18	40,66	28	67,11	38	119,51	48	266,69
9	31,58	19	43,48	29	67,67	39	129,56	49	293,33
10	32,70	20	49,75	30	68,99	40	136,12	50	439,09

Tabelle A2.6: Datensatz mit $n = 100$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,5B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	6,22	21	25,90	41	53,51	61	88,24	81	148,48
2	8,52	22	27,39	42	55,46	62	88,85	82	153,03
3	10,35	23	28,05	43	55,77	63	94,02	83	153,22
4	10,36	24	28,54	44	57,57	64	95,77	84	156,21
5	11,15	25	28,94	45	62,67	65	95,79	85	167,45
6	11,81	26	29,01	46	64,47	66	96,33	86	172,56
7	13,51	27	30,56	47	65,88	67	96,78	87	173,86
8	14,16	28	30,82	48	67,38	68	97,66	88	183,58
9	14,82	29	31,43	49	67,64	69	103,55	89	190,55
10	15,50	30	34,16	50	68,67	70	106,59	90	192,20
11	15,64	31	36,61	51	69,05	71	107,51	91	206,79
12	16,72	32	36,91	52	69,60	72	109,98	92	218,88
13	17,66	33	36,98	53	72,03	73	114,49	93	244,06
14	17,79	34	39,74	54	75,38	74	114,66	94	251,71
15	18,83	35	41,84	55	78,83	75	122,07	95	264,37
16	20,95	36	42,67	56	79,14	76	123,20	96	270,64
17	20,99	37	46,93	57	82,36	77	123,42	97	298,72
18	21,29	38	50,17	58	85,48	78	131,34	98	305,71
19	21,65	39	50,65	59	86,34	79	135,71	99	308,13
20	22,09	40	51,29	60	88,08	80	145,10	100	445,20

Tabelle A2.7: Datensatz mit $n = 20$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,9B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	14,68	5	53,24	9	70,94	13	119,62	17	207,08
2	18,64	6	53,42	10	72,27	14	127,56	18	298,93
3	22,94	7	57,44	11	85,93	15	135,06	19	349,51
4	33,55	8	59,06	12	94,87	16	189,46	20	368,23

Tabelle A2.8: Datensatz mit $n = 50$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,9B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	9,61	11	36,86	21	55,58	31	96,01	41	188,17
2	12,87	12	39,09	22	55,85	32	103,24	42	204,31
3	16,35	13	40,10	23	56,38	33	104,63	43	207,56
4	17,48	14	40,18	24	65,37	34	108,56	44	210,44
5	24,52	15	41,44	25	67,31	35	110,15	45	236,90
6	30,36	16	41,66	26	75,52	36	119,11	46	281,71
7	30,81	17	41,92	27	78,04	37	126,28	47	287,36
8	35,00	18	48,18	28	88,13	38	132,71	48	304,89
9	35,70	19	50,15	29	88,48	39	141,83	49	311,04
10	36,31	20	53,50	30	93,00	40	186,37	50	488,91

Tabelle A2.9: Datensatz mit $n = 100$ für Weibullverteilung $b = 1$, $T = 100$, $t_0 = 0,9B_{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	12,71	21	32,68	41	55,60	61	100,76	81	167,88
2	13,15	22	33,54	42	58,31	62	101,10	82	171,42
3	13,19	23	33,94	43	65,07	63	102,69	83	174,90
4	14,02	24	34,74	44	65,91	64	105,99	84	178,14
5	17,15	25	35,68	45	67,33	65	111,27	85	178,96
6	17,38	26	35,87	46	68,07	66	112,69	86	188,32
7	18,74	27	41,51	47	69,14	67	114,39	87	189,92
8	20,67	28	43,68	48	70,91	68	116,00	88	198,23
9	21,19	29	45,00	49	73,01	69	117,06	89	201,75
10	21,91	30	46,97	50	74,55	70	118,87	90	204,45
11	25,15	31	47,96	51	80,50	71	125,88	91	206,31
12	25,84	32	49,12	52	81,02	72	126,65	92	222,19
13	26,33	33	50,02	53	81,05	73	135,61	93	225,89
14	26,41	34	50,04	54	85,99	74	137,57	94	237,86
15	26,51	35	50,50	55	86,06	75	139,99	95	243,41
16	27,30	36	51,33	56	88,50	76	144,53	96	246,54
17	29,26	37	52,27	57	92,65	77	151,42	97	302,95
18	29,58	38	53,05	58	96,42	78	153,09	98	317,72
19	30,49	39	53,19	59	96,87	79	160,25	99	396,52
20	31,29	40	53,22	60	100,33	80	162,97	100	414,22

Tabelle A2.10: Datensatz mit $n = 50$ für Mischverteilung Typ A

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,0034	11	0,0075	21	0,0107	31	0,0149	41	0,7582
2	0,0035	12	0,0077	22	0,0107	32	0,0150	42	0,9603
3	0,0038	13	0,0086	23	0,0107	33	0,0168	43	0,9721
4	0,0053	14	0,0086	24	0,0110	34	0,0432	44	1,0117
5	0,0055	15	0,0090	25	0,0114	35	0,0514	45	1,1365
6	0,0060	16	0,0094	26	0,0120	36	0,0734	46	1,1470
7	0,0065	17	0,0103	27	0,0126	37	0,1589	47	1,2412
8	0,0067	18	0,0105	28	0,0128	38	0,3059	48	2,0953
9	0,0069	19	0,0106	29	0,0129	39	0,3866	49	2,6442
10	0,0074	20	0,0106	30	0,0138	40	0,4333	50	3,2174

Tabelle A2.11: Datensatz mit $n = 50$ für Mischverteilung Typ B

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,043	11	0,147	21	0,179	31	0,257	41	0,366
2	0,066	12	0,159	22	0,183	32	0,259	42	0,449
3	0,114	13	0,161	23	0,184	33	0,266	43	0,520
4	0,115	14	0,161	24	0,186	34	0,286	44	0,542
5	0,119	15	0,165	25	0,187	35	0,295	45	0,589
6	0,128	16	0,168	26	0,187	36	0,301	46	0,776
7	0,129	17	0,172	27	0,204	37	0,301	47	0,953
8	0,133	18	0,175	28	0,223	38	0,318	48	0,983
9	0,140	19	0,176	29	0,224	39	0,318	49	1,778
10	0,145	20	0,177	30	0,252	40	0,364	50	3,388

Tabelle A2.12: Datensatz mit $n = 50$ für Mischverteilung Typ C

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,066	11	0,419	21	0,708	31	0,987	41	1,243
2	0,144	12	0,483	22	0,713	32	1,022	42	1,292
3	0,161	13	0,546	23	0,750	33	1,041	43	1,335
4	0,184	14	0,582	24	0,755	34	1,050	44	1,389
5	0,268	15	0,596	25	0,804	35	1,064	45	1,409
6	0,295	16	0,613	26	0,846	36	1,100	46	1,440
7	0,358	17	0,674	27	0,861	37	1,121	47	1,616
8	0,359	18	0,686	28	0,905	38	1,128	48	2,109
9	0,391	19	0,693	29	0,949	39	1,215	49	2,406
10	0,393	20	0,699	30	0,951	40	1,236	50	2,477

Tabelle A2.13: Datensatz mit $n = 50$ für Mischverteilung Typ D

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,082	11	0,951	21	2,534	31	3,503	41	4,685
2	0,156	12	1,056	22	2,574	32	3,531	42	4,882
3	0,253	13	1,229	23	2,680	33	3,747	43	4,883
4	0,456	14	1,603	24	2,728	34	3,778	44	5,011
5	0,525	15	1,911	25	3,105	35	3,894	45	5,094
6	0,541	16	2,029	26	3,244	36	4,008	46	5,168
7	0,578	17	2,170	27	3,382	37	4,089	47	5,635
8	0,706	18	2,311	28	3,398	38	4,352	48	6,095
9	0,742	19	2,393	29	3,459	39	4,364	49	6,153
10	0,853	20	2,454	30	3,464	40	4,445	50	7,817

Tabelle A2.14: Datensatz mit $n = 50$ für Mischverteilung Typ E

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,034	11	1,055	21	11,218	31	16,858	41	21,467
2	0,119	12	1,245	22	11,606	32	16,912	42	22,134
3	0,192	13	1,292	23	13,573	33	17,042	43	23,024
4	0,224	14	1,625	24	13,632	34	17,594	44	23,388
5	0,325	15	2,068	25	14,530	35	17,867	45	23,473
6	0,385	16	2,813	26	15,407	36	17,947	46	24,764
7	0,424	17	8,731	27	15,865	37	18,538	47	25,870
8	0,477	18	9,793	28	16,057	38	19,225	48	26,187
9	0,754	19	9,931	29	16,075	39	20,767	49	27,990
10	0,781	20	10,110	30	16,141	40	20,985	50	32,491

Tabelle A2.15: Datensatz mit $n = 50$ für Mischverteilung Typ F

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,044	11	0,973	21	65,245	31	81,692	41	102,188
2	0,112	12	0,988	22	65,842	32	82,747	42	106,686
3	0,122	13	1,233	23	66,098	33	86,474	43	113,571
4	0,293	14	1,310	24	67,486	34	87,107	44	116,767
5	0,352	15	2,011	25	68,000	35	87,568	45	120,451
6	0,374	16	3,459	26	68,040	36	89,928	46	130,128
7	0,427	17	30,886	27	69,363	37	91,061	47	132,263
8	0,470	18	43,757	28	77,603	38	93,521	48	141,836
9	0,635	19	48,712	29	78,300	39	98,679	49	147,320
10	0,663	20	58,415	30	78,423	40	100,094	50	151,097

Tabelle A2.16: Datensatz mit $n = 100$ für Mischverteilung Typ A

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,0018	21	0,0064	41	0,0090	61	0,0113	81	0,3773
2	0,0028	22	0,0065	42	0,0091	62	0,0114	82	0,3800
3	0,0040	23	0,0067	43	0,0093	63	0,0115	83	0,4058
4	0,0042	24	0,0067	44	0,0094	64	0,0119	84	0,4960
5	0,0044	25	0,0067	45	0,0095	65	0,0120	85	0,4977
6	0,0048	26	0,0068	46	0,0095	66	0,0124	86	0,5175
7	0,0049	27	0,0073	47	0,0096	67	0,0129	87	0,5437
8	0,0049	28	0,0074	48	0,0098	68	0,0130	88	0,7136
9	0,0050	29	0,0075	49	0,0099	69	0,0136	89	0,8263
10	0,0050	30	0,0075	50	0,0100	70	0,0137	90	0,8308
11	0,0053	31	0,0075	51	0,0100	71	0,0138	91	1,0992
12	0,0055	32	0,0077	52	0,0101	72	0,0154	92	1,1261
13	0,0055	33	0,0078	53	0,0101	73	0,0156	93	1,2112
14	0,0058	34	0,0078	54	0,0105	74	0,1118	94	1,2418
15	0,0059	35	0,0082	55	0,0106	75	0,1613	95	1,4466
16	0,0060	36	0,0083	56	0,0106	76	0,1750	96	1,6442
17	0,0061	37	0,0083	57	0,0109	77	0,2873	97	1,6926
18	0,0061	38	0,0087	58	0,0109	78	0,2909	98	1,9227
19	0,0062	39	0,0088	59	0,0109	79	0,3112	99	2,0635
20	0,0064	40	0,0089	60	0,0112	80	0,3113	100	4,3517

Tabelle A2.17: Datensatz mit $n = 100$ für Mischverteilung Typ B

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,0017	21	0,1190	41	0,1984	61	0,2319	81	0,6615
2	0,0395	22	0,1223	42	0,1995	62	0,2338	82	0,6979
3	0,0564	23	0,1256	43	0,1996	63	0,2354	83	0,7133
4	0,0651	24	0,1258	44	0,2003	64	0,2367	84	0,7546
5	0,0699	25	0,1266	45	0,2005	65	0,2385	85	0,7653
6	0,0716	26	0,1302	46	0,2041	66	0,2402	86	0,7817
7	0,0734	27	0,1304	47	0,2077	67	0,2476	87	0,8018
8	0,0783	28	0,1310	48	0,2093	68	0,2497	88	0,8669
9	0,0878	29	0,1313	49	0,2097	69	0,2547	89	0,9081
10	0,0885	30	0,1362	50	0,2104	70	0,2620	90	1,0102
11	0,0914	31	0,1411	51	0,2140	71	0,2627	91	1,1072
12	0,0980	32	0,1470	52	0,2162	72	0,2803	92	1,4315
13	0,1003	33	0,1498	53	0,2174	73	0,2810	93	1,4497
14	0,1038	34	0,1687	54	0,2198	74	0,2866	94	1,4726
15	0,1046	35	0,1770	55	0,2201	75	0,2872	95	1,5203
16	0,1049	36	0,1812	56	0,2216	76	0,3178	96	1,7513
17	0,1092	37	0,1877	57	0,2220	77	0,3507	97	2,8304
18	0,1112	38	0,1958	58	0,2224	78	0,4128	98	2,9564
19	0,1127	39	0,1970	59	0,2233	79	0,4856	99	3,2979
20	0,1130	40	0,1982	60	0,2312	80	0,6594	100	3,3962

Tabelle A2.18: Datensatz mit $n = 100$ für Mischverteilung Typ C

Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe
1	0,033	21	0,545	41	0,749	61	0,996	81	1,244
2	0,109	22	0,553	42	0,765	62	1,016	82	1,278
3	0,145	23	0,559	43	0,766	63	1,033	83	1,285
4	0,150	24	0,597	44	0,785	64	1,035	84	1,289
5	0,161	25	0,613	45	0,806	65	1,048	85	1,294
6	0,184	26	0,647	46	0,810	66	1,073	86	1,306
7	0,307	27	0,651	47	0,834	67	1,074	87	1,417
8	0,307	28	0,663	48	0,842	68	1,086	88	1,443
9	0,312	29	0,669	49	0,847	69	1,095	89	1,448
10	0,327	30	0,672	50	0,849	70	1,098	90	1,488
11	0,365	31	0,682	51	0,859	71	1,114	91	1,540
12	0,366	32	0,687	52	0,868	72	1,119	92	1,566
13	0,391	33	0,691	53	0,910	73	1,120	93	1,597
14	0,463	34	0,694	54	0,913	74	1,132	94	1,649
15	0,466	35	0,700	55	0,918	75	1,147	95	1,662
16	0,471	36	0,703	56	0,954	76	1,156	96	1,896
17	0,489	37	0,707	57	0,985	77	1,172	97	2,346
18	0,515	38	0,727	58	0,986	78	1,178	98	2,398
19	0,526	39	0,740	59	0,991	79	1,180	99	2,665
20	0,536	40	0,741	60	0,991	80	1,196	100	4,812

Tabelle A2.19: Datensatz mit $n = 100$ für Mischverteilung Typ D

Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe
1	0,031	21	1,109	41	2,531	61	3,414	81	4,527
2	0,049	22	1,402	42	2,556	62	3,473	82	4,638
3	0,067	23	1,407	43	2,572	63	3,477	83	4,655
4	0,138	24	1,432	44	2,572	64	3,525	84	4,720
5	0,232	25	1,456	45	2,622	65	3,527	85	4,800
6	0,323	26	1,496	46	2,653	66	3,588	86	4,822
7	0,330	27	1,702	47	2,786	67	3,596	87	4,827
8	0,380	28	1,704	48	2,789	68	3,606	88	4,876
9	0,449	29	1,763	49	2,800	69	3,610	89	5,026
10	0,503	30	1,800	50	2,805	70	3,621	90	5,072
11	0,556	31	1,808	51	2,813	71	3,623	91	5,074
12	0,562	32	1,910	52	3,052	72	3,766	92	5,116
13	0,573	33	2,027	53	3,105	73	3,899	93	5,311
14	0,580	34	2,036	54	3,196	74	3,971	94	5,312
15	0,621	35	2,098	55	3,208	75	4,006	95	5,834
16	0,710	36	2,203	56	3,222	76	4,091	96	6,087
17	0,723	37	2,235	57	3,259	77	4,143	97	6,152
18	0,729	38	2,379	58	3,315	78	4,204	98	6,293
19	0,800	39	2,408	59	3,371	79	4,348	99	6,313
20	1,106	40	2,456	60	3,389	80	4,416	100	6,644

Tabelle A2.20: Datensatz mit $n = 100$ für Mischverteilung Typ E

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,021	21	1,635	41	10,616	61	15,182	81	20,432
2	0,078	22	1,646	42	10,815	62	15,509	82	20,455
3	0,095	23	1,810	43	11,821	63	15,613	83	22,362
4	0,115	24	1,997	44	12,031	64	15,660	84	23,445
5	0,135	25	2,026	45	12,103	65	15,952	85	23,496
6	0,174	26	2,240	46	12,219	66	16,339	86	23,514
7	0,198	27	2,263	47	12,225	67	16,409	87	24,020
8	0,198	28	2,527	48	12,479	68	16,729	88	24,326
9	0,204	29	2,760	49	12,537	69	16,981	89	24,696
10	0,256	30	3,068	50	12,646	70	17,459	90	24,783
11	0,302	31	3,215	51	12,932	71	17,563	91	24,869
12	0,380	32	3,220	52	13,228	72	17,753	92	25,669
13	0,457	33	4,465	53	13,331	73	17,836	93	26,788
14	0,580	34	4,911	54	13,420	74	18,442	94	26,833
15	0,634	35	5,340	55	13,621	75	18,940	95	26,878
16	0,830	36	7,221	56	13,757	76	19,698	96	27,795
17	0,947	37	8,778	57	13,822	77	19,891	97	27,945
18	1,134	38	9,367	58	14,274	78	19,997	98	30,761
19	1,150	39	9,512	59	14,450	79	20,163	99	32,327
20	1,618	40	9,666	60	15,060	80	20,270	100	33,395

Tabelle A2.21: Datensatz mit $n = 100$ für Mischverteilung Typ F

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,016	21	1,049	41	55,654	61	87,418	81	107,646
2	0,091	22	1,744	42	55,689	62	88,505	82	111,566
3	0,094	23	1,785	43	58,489	63	91,073	83	112,242
4	0,117	24	1,867	44	58,560	64	91,980	84	112,887
5	0,124	25	1,869	45	59,426	65	95,226	85	113,218
6	0,131	26	1,968	46	59,865	66	96,047	86	114,369
7	0,192	27	2,014	47	61,831	67	96,709	87	114,384
8	0,294	28	2,742	48	66,867	68	97,942	88	120,978
9	0,331	29	3,376	49	68,450	69	98,271	89	121,140
10	0,373	30	3,898	50	69,026	70	98,919	90	124,842
11	0,445	31	24,700	51	69,482	71	98,947	91	130,713
12	0,464	32	34,469	52	70,482	72	101,440	92	131,896
13	0,477	33	38,242	53	73,943	73	101,842	93	132,118
14	0,553	34	43,319	54	74,157	74	103,048	94	134,167
15	0,647	35	44,305	55	76,224	75	103,654	95	135,961
16	0,679	36	48,746	56	79,762	76	104,055	96	136,060
17	0,680	37	49,727	57	82,302	77	105,094	97	144,088
18	0,709	38	52,852	58	82,466	78	105,465	98	146,124
19	0,728	39	54,661	59	82,919	79	105,944	99	148,997
20	0,746	40	54,840	60	85,329	80	107,498	100	151,693

Tabelle A2.22: Datensatz mit $n = 100$ für Mischverteilung Typ AB

Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe
1	0,0021	21	0,0099	41	0,0895	61	0,1962	81	0,3673
2	0,0041	22	0,0101	42	0,0933	62	0,1996	82	0,3971
3	0,0052	23	0,0103	43	0,0991	63	0,2133	83	0,4686
4	0,0064	24	0,0105	44	0,1051	64	0,2166	84	0,4812
5	0,0066	25	0,0107	45	0,1262	65	0,2168	85	0,4926
6	0,0066	26	0,0116	46	0,1309	66	0,2187	86	0,4963
7	0,0066	27	0,0119	47	0,1321	67	0,2199	87	0,5498
8	0,0069	28	0,0119	48	0,1387	68	0,2274	88	0,5683
9	0,0070	29	0,0125	49	0,1464	69	0,2274	89	0,5813
10	0,0074	30	0,0129	50	0,1484	70	0,2439	90	0,9162
11	0,0076	31	0,0132	51	0,1491	71	0,2654	91	1,1084
12	0,0078	32	0,0138	52	0,1526	72	0,2659	92	1,1379
13	0,0080	33	0,0138	53	0,1538	73	0,2692	93	1,2624
14	0,0083	34	0,0153	54	0,1706	74	0,2776	94	1,2865
15	0,0083	35	0,0156	55	0,1738	75	0,2793	95	1,4546
16	0,0089	36	0,0163	56	0,1764	76	0,2953	96	1,8027
17	0,0090	37	0,0362	57	0,1802	77	0,2984	97	1,8714
18	0,0093	38	0,0543	58	0,1904	78	0,3033	98	2,1945
19	0,0097	39	0,0548	59	0,1911	79	0,3134	99	2,9002
20	0,0097	40	0,0564	60	0,1939	80	0,3436	100	3,3015

Tabelle A2.23: Datensatz mit $n = 100$ für Mischverteilung Typ BC

Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe
1	0,028	21	0,179	41	0,264	61	0,700	81	1,027
2	0,049	22	0,179	42	0,272	62	0,709	82	1,070
3	0,054	23	0,192	43	0,277	63	0,722	83	1,098
4	0,065	24	0,193	44	0,280	64	0,755	84	1,121
5	0,069	25	0,193	45	0,296	65	0,780	85	1,122
6	0,080	26	0,194	46	0,351	66	0,786	86	1,129
7	0,081	27	0,199	47	0,427	67	0,793	87	1,139
8	0,103	28	0,200	48	0,464	68	0,795	88	1,151
9	0,104	29	0,202	49	0,484	69	0,839	89	1,203
10	0,109	30	0,202	50	0,521	70	0,857	90	1,209
11	0,110	31	0,205	51	0,538	71	0,861	91	1,279
12	0,122	32	0,215	52	0,628	72	0,866	92	1,285
13	0,127	33	0,219	53	0,646	73	0,928	93	1,416
14	0,139	34	0,219	54	0,652	74	0,940	94	1,421
15	0,143	35	0,221	55	0,653	75	0,957	95	1,601
16	0,148	36	0,223	56	0,653	76	0,976	96	1,841
17	0,150	37	0,228	57	0,655	77	0,988	97	1,850
18	0,162	38	0,230	58	0,682	78	1,012	98	2,017
19	0,164	39	0,231	59	0,682	79	1,020	99	2,098
20	0,177	40	0,239	60	0,695	80	1,026	100	3,255

Tabelle A2.24: Datensatz mit $n = 100$ für Mischverteilung Typ EF

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,016	21	1,456	41	15,395	61	41,192	81	87,880
2	0,051	22	1,785	42	17,140	62	41,427	82	94,020
3	0,062	23	1,862	43	17,243	63	46,640	83	98,424
4	0,099	24	2,379	44	18,121	64	53,388	84	100,844
5	0,131	25	2,923	45	18,218	65	53,632	85	102,988
6	0,240	26	2,944	46	18,291	66	55,648	86	112,840
7	0,307	27	3,006	47	19,279	67	61,009	87	112,905
8	0,377	28	3,030	48	20,156	68	64,335	88	113,481
9	0,409	29	5,620	49	21,063	69	68,397	89	116,969
10	0,462	30	6,530	50	21,257	70	69,614	90	119,148
11	0,473	31	7,223	51	22,226	71	70,741	91	121,211
12	0,506	32	7,954	52	22,745	72	75,333	92	122,229
13	0,565	33	9,270	53	23,772	73	77,135	93	129,258
14	0,582	34	9,590	54	24,695	74	78,239	94	129,880
15	0,594	35	11,681	55	25,913	75	80,714	95	133,239
16	0,778	36	11,968	56	27,753	76	81,404	96	137,048
17	0,783	37	12,427	57	28,244	77	82,758	97	141,283
18	0,848	38	13,889	58	30,431	78	82,813	98	153,098
19	0,987	39	13,977	59	30,617	79	83,700	99	155,718
20	1,097	40	14,668	60	32,429	80	84,570	100	178,120

Tabelle A2.25: Datensatz mit $n = 100$ für Mischverteilung Typ AF

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,0028	21	0,0094	41	0,0422	61	1,394	81	70,202
2	0,0048	22	0,0097	42	0,0513	62	1,777	82	71,792
3	0,0050	23	0,0099	43	0,0819	63	1,813	83	73,381
4	0,0053	24	0,0104	44	0,1020	64	1,826	84	74,995
5	0,0054	25	0,0104	45	0,1307	65	1,883	85	78,026
6	0,0057	26	0,0106	46	0,2177	66	1,959	86	79,737
7	0,0064	27	0,0109	47	0,4180	67	2,286	87	82,231
8	0,0066	28	0,0110	48	0,4252	68	2,301	88	86,207
9	0,0066	29	0,0119	49	0,4555	69	2,707	89	89,373
10	0,0068	30	0,0120	50	0,4787	70	2,729	90	109,646
11	0,0071	31	0,0124	51	0,5245	71	3,164	91	110,505
12	0,0072	32	0,0125	52	0,5587	72	3,903	92	111,183
13	0,0073	33	0,0126	53	0,5734	73	5,119	93	111,398
14	0,0076	34	0,0129	54	0,7497	74	27,359	94	111,739
15	0,0079	35	0,0137	55	0,9248	75	44,015	95	112,031
16	0,0086	36	0,0140	56	0,9909	76	47,775	96	113,958
17	0,0088	37	0,0144	57	1,112	77	55,417	97	117,395
18	0,0089	38	0,0151	58	1,128	78	62,106	98	119,063
19	0,0093	39	0,0153	59	1,269	79	63,904	99	121,957
20	0,0093	40	0,0353	60	1,290	80	70,168	100	145,494

Tabelle A2.26: Datensatz mit $n = 100$ für Mischverteilung Typ ADF

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,0022	21	0,0124	41	0,7364	61	3,465	81	73,372
2	0,0046	22	0,0135	42	0,8035	62	3,598	82	76,778
3	0,0056	23	0,0135	43	0,8544	63	3,737	83	78,594
4	0,0058	24	0,0139	44	1,171	64	3,761	84	81,498
5	0,0065	25	0,0140	45	1,232	65	4,438	85	83,493
6	0,0070	26	0,0719	46	1,271	66	4,516	86	87,440
7	0,0073	27	0,0762	47	1,417	67	4,598	87	88,445
8	0,0073	28	0,0775	48	1,538	68	4,708	88	94,969
9	0,0077	29	0,1604	49	1,599	69	4,921	89	97,673
10	0,0082	30	0,1978	50	1,702	70	5,018	90	98,147
11	0,0097	31	0,2070	51	2,071	71	5,136	91	99,557
12	0,0099	32	0,2172	52	2,153	72	5,176	92	115,284
13	0,0100	33	0,2283	53	2,529	73	5,679	93	119,209
14	0,0104	34	0,2882	54	2,637	74	21,247	94	125,022
15	0,0105	35	0,2942	55	2,975	75	30,686	95	126,472
16	0,0105	36	0,3071	56	2,977	76	46,093	96	133,359
17	0,0110	37	0,3806	57	2,991	77	54,559	97	136,329
18	0,0116	38	0,5378	58	3,086	78	64,776	98	152,412
19	0,0118	39	0,5779	59	3,300	79	70,566	99	156,935
20	0,0119	40	0,6132	60	3,434	80	71,601	100	158,461

Tabelle A2.27: Datensatz mit $n = 100$ für Mischverteilung Typ BDE

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,024	21	0,203	41	0,578	61	3,078	81	12,544
2	0,036	22	0,209	42	0,673	62	3,257	82	13,673
3	0,063	23	0,210	43	0,827	63	3,283	83	14,074
4	0,072	24	0,217	44	0,841	64	3,370	84	14,694
5	0,084	25	0,231	45	0,890	65	3,386	85	15,736
6	0,089	26	0,232	46	0,898	66	3,433	86	16,126
7	0,120	27	0,241	47	0,987	67	3,641	87	16,440
8	0,121	28	0,243	48	1,024	68	3,836	88	16,740
9	0,133	29	0,251	49	1,154	69	4,007	89	17,416
10	0,139	30	0,278	50	1,315	70	4,024	90	18,215
11	0,144	31	0,289	51	1,416	71	4,278	91	19,141
12	0,146	32	0,292	52	1,770	72	4,693	92	20,254
13	0,148	33	0,294	53	2,076	73	4,707	93	20,865
14	0,157	34	0,306	54	2,292	74	4,710	94	20,916
15	0,171	35	0,308	55	2,422	75	4,917	95	23,081
16	0,180	36	0,314	56	2,449	76	5,256	96	24,955
17	0,183	37	0,372	57	2,638	77	5,848	97	26,132
18	0,195	38	0,389	58	2,640	78	6,010	98	27,344
19	0,198	39	0,465	59	2,696	79	6,357	99	27,861
20	0,202	40	0,526	60	2,969	80	10,274	100	31,055

Tabelle A2.28: Datensatz mit $n = 100$ für Mischverteilung Typ CDE

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,031	21	0,765	41	1,341	61	3,497	81	10,997
2	0,109	22	0,795	42	1,426	62	3,630	82	11,473
3	0,126	23	0,897	43	1,473	63	3,717	83	11,762
4	0,186	24	0,917	44	1,546	64	3,718	84	11,928
5	0,212	25	0,923	45	1,611	65	3,754	85	12,555
6	0,239	26	0,969	46	1,672	66	3,828	86	15,152
7	0,246	27	0,976	47	1,673	67	3,913	87	15,707
8	0,351	28	0,986	48	1,957	68	4,052	88	16,016
9	0,407	29	1,028	49	1,985	69	4,218	89	16,542
10	0,411	30	1,037	50	2,007	70	4,303	90	17,421
11	0,461	31	1,061	51	2,213	71	4,380	91	17,474
12	0,498	32	1,088	52	2,342	72	4,393	92	19,597
13	0,515	33	1,105	53	2,385	73	4,446	93	23,107
14	0,518	34	1,129	54	2,500	74	4,460	94	23,261
15	0,623	35	1,200	55	2,554	75	5,566	95	23,472
16	0,655	36	1,230	56	2,816	76	5,648	96	24,014
17	0,656	37	1,258	57	2,888	77	5,726	97	24,612
18	0,664	38	1,308	58	3,084	78	6,041	98	25,079
19	0,678	39	1,333	59	3,096	79	6,047	99	25,265
20	0,718	40	1,341	60	3,167	80	6,659	100	27,746

Tabelle A2.29: Datensatz mit $n = 50$ für Mischverteilung Typ $B_{0,1}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,041	11	0,139	21	0,190	31	0,244	41	0,417
2	0,063	12	0,142	22	0,194	32	0,247	42	0,454
3	0,067	13	0,144	23	0,207	33	0,265	43	0,478
4	0,101	14	0,149	24	0,211	34	0,271	44	0,541
5	0,103	15	0,151	25	0,219	35	0,282	45	0,611
6	0,124	16	0,153	26	0,220	36	0,283	46	1,087
7	0,129	17	0,163	27	0,220	37	0,284	47	1,347
8	0,130	18	0,168	28	0,230	38	0,286	48	1,823
9	0,137	19	0,170	29	0,236	39	0,300	49	2,308
10	0,138	20	0,172	30	0,240	40	0,415	50	2,436

Tabelle A2.30: Datensatz mit $n = 50$ für Mischverteilung Typ $B_{0,5}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,084	11	0,164	21	0,205	31	0,232	41	0,544
2	0,093	12	0,168	22	0,209	32	0,236	42	0,601
3	0,114	13	0,176	23	0,212	33	0,238	43	0,835
4	0,119	14	0,177	24	0,212	34	0,242	44	0,992
5	0,126	15	0,186	25	0,216	35	0,246	45	1,158
6	0,126	16	0,187	26	0,217	36	0,251	46	1,300
7	0,136	17	0,187	27	0,219	37	0,260	47	1,340
8	0,147	18	0,189	28	0,220	38	0,262	48	1,386
9	0,149	19	0,190	29	0,223	39	0,292	49	1,881
10	0,151	20	0,201	30	0,225	40	0,330	50	3,142

Tabelle A2.31: Datensatz mit $n = 50$ für Mischverteilung Typ $B_{0,9}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,117	11	0,160	21	0,187	31	0,212	41	0,609
2	0,130	12	0,163	22	0,190	32	0,216	42	0,727
3	0,140	13	0,171	23	0,193	33	0,230	43	0,812
4	0,146	14	0,177	24	0,195	34	0,238	44	0,939
5	0,151	15	0,178	25	0,198	35	0,240	45	0,986
6	0,153	16	0,179	26	0,200	36	0,258	46	1,300
7	0,154	17	0,180	27	0,200	37	0,276	47	1,663
8	0,155	18	0,183	28	0,200	38	0,397	48	2,206
9	0,156	19	0,186	29	0,204	39	0,587	49	2,343
10	0,157	20	0,186	30	0,212	40	0,588	50	2,813

Tabelle A2.32: Datensatz mit $n = 50$ für Mischverteilung Typ $D_{0,1}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,062	11	0,789	21	2,029	31	3,149	41	4,099
2	0,113	12	0,941	22	2,033	32	3,197	42	4,107
3	0,153	13	1,039	23	2,102	33	3,550	43	4,125
4	0,188	14	1,059	24	2,196	34	3,555	44	4,327
5	0,221	15	1,375	25	2,215	35	3,579	45	4,466
6	0,338	16	1,375	26	2,329	36	3,605	46	4,556
7	0,387	17	1,505	27	2,478	37	3,688	47	4,588
8	0,436	18	1,884	28	2,646	38	3,766	48	5,248
9	0,444	19	1,897	29	2,916	39	3,774	49	5,280
10	0,550	20	2,028	30	2,965	40	3,929	50	5,512

Tabelle A2.33: Datensatz mit $n = 50$ für Mischverteilung Typ $D_{0,5}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,071	11	1,151	21	2,383	31	3,353	41	4,371
2	0,133	12	1,585	22	2,484	32	3,388	42	4,642
3	0,190	13	1,845	23	2,541	33	3,501	43	4,733
4	0,248	14	1,856	24	2,573	34	3,536	44	4,743
5	0,276	15	1,903	25	2,658	35	3,543	45	4,773
6	0,418	16	2,076	26	2,756	36	3,629	46	4,873
7	0,579	17	2,076	27	3,100	37	3,846	47	5,025
8	0,760	18	2,160	28	3,159	38	3,928	48	5,166
9	0,915	19	2,179	29	3,184	39	4,238	49	5,268
10	1,140	20	2,216	30	3,296	40	4,367	50	5,943

Tabelle A2.34: Datensatz mit $n = 50$ für Mischverteilung Typ $D_{0,9}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,123	11	0,829	21	3,243	31	3,644	41	4,441
2	0,232	12	1,035	22	3,349	32	3,766	42	4,536
3	0,337	13	1,220	23	3,388	33	3,798	43	4,572
4	0,344	14	1,285	24	3,457	34	3,810	44	4,599
5	0,350	15	1,829	25	3,510	35	3,905	45	4,694
6	0,438	16	2,759	26	3,517	36	3,918	46	4,724
7	0,585	17	2,901	27	3,540	37	3,943	47	4,848
8	0,600	18	2,949	28	3,579	38	4,189	48	5,097
9	0,771	19	3,097	29	3,605	39	4,269	49	5,670
10	0,804	20	3,122	30	3,627	40	4,380	50	5,761

Tabelle A2.35: Datensatz mit $n = 50$ für Mischverteilung Typ $F_{0,1}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,091	11	1,467	21	64,157	31	91,108	41	126,263
2	0,190	12	1,564	22	64,309	32	94,224	42	126,834
3	0,223	13	1,573	23	67,088	33	95,618	43	131,178
4	0,388	14	2,628	24	67,731	34	96,360	44	133,131
5	0,486	15	2,637	25	68,253	35	98,525	45	139,361
6	0,516	16	3,305	26	72,461	36	100,941	46	141,421
7	0,695	17	47,489	27	78,738	37	108,405	47	143,759
8	0,761	18	49,312	28	79,878	38	110,219	48	146,536
9	1,056	19	56,426	29	80,240	39	117,435	49	156,538
10	1,466	20	62,443	30	89,644	40	123,623	50	190,149

Tabelle A2.36: Datensatz mit $n = 50$ für Mischverteilung Typ $F_{0,5}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,144	11	1,724	21	66,856	31	93,638	41	111,268
2	0,209	12	2,170	22	68,990	32	94,300	42	111,697
3	0,246	13	3,105	23	71,257	33	94,976	43	117,718
4	0,530	14	3,671	24	77,854	34	96,166	44	120,289
5	0,543	15	46,309	25	81,537	35	98,895	45	123,224
6	0,594	16	46,745	26	81,663	36	102,201	46	129,463
7	0,935	17	48,692	27	83,297	37	103,877	47	131,596
8	1,077	18	57,244	28	83,510	38	103,895	48	140,480
9	1,361	19	62,897	29	91,239	39	107,955	49	142,340
10	1,438	20	64,477	30	93,195	40	108,718	50	142,433

Tabelle A2.37: Datensatz mit $n = 50$ für Mischverteilung Typ $F_{0,9}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	0,092	11	1,208	21	71,300	31	84,655	41	106,442
2	0,127	12	1,280	22	71,460	32	85,026	42	106,904
3	0,223	13	1,998	23	71,489	33	85,356	43	110,247
4	0,332	14	2,466	24	71,747	34	86,351	44	110,743
5	0,334	15	2,792	25	71,837	35	89,584	45	110,939
6	0,472	16	3,294	26	78,096	36	92,057	46	113,019
7	0,567	17	52,229	27	80,294	37	92,262	47	113,082
8	0,622	18	58,489	28	80,339	38	94,133	48	116,059
9	0,679	19	65,043	29	82,249	39	97,621	49	128,706
10	0,947	20	67,939	30	83,788	40	98,299	50	132,837

Tabelle A2.38: Datensatz mit $n = 50$ für Seriensystem Typ I_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,94	11	20,48	21	40,85	31	73,68	41	103,16
2	4,32	12	25,35	22	41,47	32	77,95	42	108,11
3	4,85	13	26,49	23	45,48	33	80,17	43	112,19
4	8,72	14	28,48	24	46,02	34	81,31	44	116,97
5	9,85	15	29,74	25	51,69	35	82,92	45	118,72
6	12,85	16	31,29	26	56,70	36	87,19	46	121,16
7	12,89	17	37,29	27	58,49	37	89,64	47	132,22
8	14,64	18	38,59	28	63,74	38	90,41	48	147,06
9	14,73	19	39,27	29	69,16	39	100,16	49	153,15
10	16,27	20	39,89	30	69,39	40	102,34	50	154,53

Tabelle A2.39: Datensatz mit $n = 50$ für Seriensystem Typ II_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,50	11	20,98	21	45,26	31	57,23	41	82,87
2	5,95	12	22,38	22	45,45	32	60,56	42	84,07
3	6,11	13	25,88	23	45,68	33	61,62	43	84,89
4	7,91	14	26,99	24	45,88	34	67,64	44	85,56
5	11,33	15	28,41	25	48,11	35	69,31	45	87,87
6	12,51	16	33,83	26	51,55	36	78,72	46	89,59
7	12,98	17	37,46	27	51,67	37	79,13	47	91,80
8	13,69	18	40,46	28	51,72	38	79,83	48	103,62
9	13,72	19	42,54	29	53,65	39	80,04	49	107,05
10	16,21	20	44,26	30	56,72	40	80,29	50	116,09

Tabelle A2.40: Datensatz mit $n = 50$ für Seriensystem Typ III_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,62	11	15,35	21	24,58	31	31,20	41	42,58
2	3,47	12	18,49	22	24,71	32	33,66	42	43,26
3	4,38	13	20,73	23	25,38	33	34,94	43	43,49
4	7,92	14	21,56	24	25,78	34	36,99	44	43,91
5	8,03	15	21,71	25	25,89	35	37,92	45	44,59
6	11,08	16	22,25	26	26,94	36	39,08	46	45,56
7	11,50	17	22,51	27	27,06	37	39,08	47	47,24
8	12,47	18	23,15	28	29,88	38	39,34	48	48,35
9	13,34	19	24,03	29	30,05	39	40,34	49	48,46
10	14,58	20	24,24	30	30,42	40	42,42	50	59,55

Tabelle A2.41: Datensatz mit $n = 50$ für Seriensystem Typ IV_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,14	11	32,46	21	64,18	31	83,57	41	112,99
2	3,37	12	36,90	22	66,45	32	84,82	42	117,19
3	9,69	13	43,19	23	68,23	33	85,94	43	118,71
4	10,16	14	43,72	24	69,55	34	86,08	44	121,93
5	11,40	15	47,65	25	72,12	35	90,74	45	123,70
6	13,63	16	50,80	26	73,85	36	97,34	46	126,88
7	16,53	17	53,86	27	75,72	37	101,22	47	127,85
8	17,16	18	53,92	28	76,13	38	102,61	48	129,70
9	28,37	19	55,08	29	79,21	39	104,50	49	131,05
10	30,01	20	56,91	30	79,73	40	112,08	50	131,22

Tabelle A2.42: Datensatz mit $n = 50$ für Seriensystem Typ V_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,46	11	25,51	21	46,66	31	65,03	41	75,06
2	2,85	12	31,62	22	48,53	32	66,10	42	77,70
3	5,32	13	32,64	23	52,47	33	67,37	43	79,94
4	13,39	14	32,89	24	53,64	34	69,79	44	81,14
5	13,79	15	36,58	25	54,94	35	71,78	45	81,34
6	14,44	16	37,61	26	55,84	36	71,78	46	81,43
7	14,53	17	39,00	27	57,78	37	71,93	47	83,98
8	16,18	18	40,42	28	61,20	38	72,52	48	86,59
9	18,01	19	45,20	29	64,82	39	73,70	49	88,04
10	21,01	20	45,87	30	64,97	40	74,71	50	92,41

Tabelle A2.43: Datensatz mit $n = 50$ für Seriensystem Typ VI_s

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	1,96	11	17,47	21	25,75	31	30,16	41	36,29
2	2,81	12	20,32	22	26,01	32	31,02	42	36,64
3	4,92	13	21,41	23	26,29	33	31,05	43	36,75
4	5,64	14	22,44	24	26,35	34	31,07	44	37,03
5	6,93	15	22,62	25	26,95	35	31,81	45	37,97
6	7,04	16	22,78	26	27,15	36	32,50	46	40,04
7	11,00	17	23,22	27	27,87	37	32,99	47	41,27
8	15,19	18	23,79	28	28,10	38	33,05	48	41,53
9	15,56	19	24,79	29	28,12	39	34,47	49	41,98
10	17,18	20	25,27	30	28,17	40	34,78	50	43,85

Tabelle A2.44: Datensatz mit $n = 50$ für Seriensystem Typ $II_{s,0,1}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	2,52	11	28,15	21	47,12	31	60,81	41	82,00
2	7,28	12	32,10	22	50,37	32	61,37	42	83,43
3	7,99	13	32,18	23	50,65	33	63,08	43	84,84
4	10,70	14	32,63	24	52,84	34	63,76	44	85,38
5	10,88	15	33,04	25	58,64	35	66,75	45	85,50
6	12,66	16	33,11	26	59,70	36	67,39	46	86,08
7	12,68	17	33,80	27	59,72	37	72,02	47	90,03
8	18,58	18	35,69	28	59,88	38	72,52	48	90,19
9	19,07	19	38,24	29	59,97	39	72,87	49	92,11
10	21,04	20	38,39	30	60,20	40	77,56	50	102,63

Tabelle A2.45: Datensatz mit $n = 50$ für Seriensystem Typ $II_{s,0,5}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	6,47	11	31,27	21	44,62	31	62,35	41	76,00
2	7,73	12	31,41	22	44,70	32	62,70	42	76,46
3	10,34	13	33,30	23	46,12	33	63,35	43	78,85
4	12,45	14	37,15	24	46,14	34	64,09	44	81,37
5	13,94	15	37,18	25	47,77	35	66,55	45	82,71
6	17,88	16	38,15	26	47,84	36	68,77	46	84,90
7	19,25	17	38,64	27	54,02	37	69,33	47	86,78
8	20,66	18	42,42	28	54,28	38	70,21	48	87,31
9	23,27	19	42,50	29	55,85	39	70,28	49	99,42
10	28,26	20	43,18	30	57,96	40	72,80	50	99,58

Tabelle A2.46: Datensatz mit $n = 50$ für Seriensystem Typ $V_{s,0,1}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	2,00	11	22,11	21	45,74	31	62,43	41	73,13
2	3,54	12	23,36	22	47,92	32	63,15	42	73,23
3	7,63	13	25,22	23	48,94	33	66,40	43	73,97
4	7,64	14	29,41	24	52,18	34	67,77	44	74,13
5	10,01	15	35,28	25	53,41	35	67,95	45	77,13
6	14,61	16	36,57	26	55,32	36	69,21	46	79,18
7	15,99	17	36,92	27	55,84	37	69,53	47	79,20
8	16,25	18	37,26	28	58,62	38	70,42	48	79,24
9	18,39	19	39,76	29	59,14	39	71,21	49	84,51
10	21,56	20	44,71	30	60,75	40	72,14	50	85,44

Tabelle A2.47: Datensatz mit $n = 50$ für Seriensystem Typ $V_{s,0,5}$ ¹⁰

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	6,57	11	32,21	21	51,20	31	63,11	41	74,75
2	7,22	12	32,89	22	51,79	32	67,31	42	75,69
3	11,50	13	38,92	23	54,80	33	67,47	43	77,89
4	12,76	14	39,61	24	55,68	34	67,52	44	79,00
5	13,66	15	43,57	25	55,76	35	68,50	45	79,72
6	17,99	16	44,47	26	59,04	36	68,84	46	79,77
7	18,73	17	48,59	27	59,20	37	69,42	47	82,39
8	23,17	18	49,79	28	61,47	38	72,35	48	84,74
9	24,58	19	50,41	29	61,52	39	73,11	49	88,21
10	32,09	20	50,55	30	62,51	40	73,62	50	89,22

Tabelle A2.48: Datensatz mit $n = 50$ für Seriensystem Typ I_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	33,50	11	91,69	21	130,00	31	199,45	41	371,08
2	53,86	12	93,14	22	140,40	32	207,93	42	425,22
3	67,52	13	97,56	23	146,56	33	242,61	43	458,44
4	76,48	14	100,15	24	147,16	34	263,49	44	510,14
5	78,26	15	102,93	25	151,55	35	267,02	45	526,21
6	78,89	16	103,18	26	164,70	36	278,66	46	538,96
7	79,78	17	103,83	27	185,43	37	298,23	47	675,73
8	83,96	18	106,44	28	190,19	38	303,49	48	740,69
9	89,12	19	123,91	29	192,26	39	337,47	49	794,61
10	91,03	20	128,70	30	196,81	40	354,51	50	836,99

Tabelle A2.49: Datensatz mit $n = 50$ für Seriensystem Typ II_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	37,04	11	94,14	21	121,44	31	144,48	41	215,44
2	53,27	12	96,49	22	121,98	32	144,97	42	233,24
3	58,46	13	96,91	23	123,02	33	151,80	43	251,70
4	73,82	14	98,52	24	124,93	34	153,32	44	257,75
5	75,29	15	100,00	25	125,00	35	156,59	45	263,31
6	76,07	16	108,09	26	126,41	36	158,87	46	293,92
7	83,53	17	111,73	27	128,90	37	166,72	47	303,50
8	85,51	18	113,01	28	133,57	38	170,24	48	360,64
9	88,29	19	113,90	29	138,46	39	171,99	49	529,66
10	92,25	20	114,40	30	141,80	40	199,04	50	592,94

Tabelle A2.50: Datensatz mit $n = 50$ für Seriensystem Typ III_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	25,14	11	73,63	21	89,26	31	116,24	41	140,50
2	46,04	12	74,33	22	96,26	32	122,38	42	146,11
3	49,04	13	75,23	23	97,36	33	125,25	43	150,09
4	58,60	14	78,88	24	101,32	34	127,34	44	161,81
5	59,23	15	83,50	25	103,45	35	127,88	45	187,60
6	65,04	16	83,70	26	107,49	36	129,92	46	202,48
7	65,87	17	84,63	27	110,42	37	130,26	47	236,68
8	71,62	18	85,44	28	114,30	38	131,16	48	243,22
9	72,78	19	87,07	29	114,74	39	137,02	49	256,30
10	73,54	20	87,39	30	115,08	40	140,35	50	376,12

Tabelle A2.51: Datensatz mit $n = 50$ für Seriensystem Typ IV_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	41,65	11	99,39	21	145,21	31	438,83	41	1604,8
2	53,52	12	102,43	22	201,30	32	478,23	42	1820,2
3	64,39	13	107,51	23	210,90	33	497,96	43	1889,8
4	69,16	14	112,07	24	219,86	34	501,51	44	2205,0
5	75,61	15	113,22	25	227,33	35	507,04	45	2215,4
6	76,31	16	115,36	26	278,35	36	717,85	46	2346,5
7	78,60	17	124,95	27	309,65	37	851,02	47	2359,2
8	85,51	18	128,41	28	368,12	38	943,31	48	3269,4
9	90,84	19	134,66	29	381,81	39	1112,1	49	3408,8
10	94,41	20	136,46	30	405,37	40	1470,3	50	5800,5

Tabelle A2.52: Datensatz mit $n = 50$ für Seriensystem Typ V_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	40,15	11	77,62	21	109,28	31	145,84	41	382,22
2	53,20	12	81,42	22	109,90	32	158,19	42	421,11
3	54,72	13	86,44	23	110,51	33	174,85	43	427,57
4	65,23	14	88,93	24	111,48	34	181,81	44	464,48
5	66,27	15	93,20	25	114,37	35	199,68	45	473,88
6	68,10	16	96,45	26	115,42	36	231,48	46	564,31
7	68,58	17	99,78	27	124,43	37	242,36	47	735,28
8	71,17	18	105,06	28	124,56	38	243,27	48	871,96
9	72,57	19	105,46	29	130,12	39	284,33	49	1048,5
10	76,57	20	108,24	30	143,26	40	357,53	50	1080,0

Tabelle A2.53: Datensatz mit $n = 50$ für Seriensystem Typ VI_p

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	35,18	11	84,72	21	103,39	31	126,13	41	200,82
2	41,27	12	85,71	22	105,32	32	130,63	42	201,42
3	50,28	13	88,23	23	106,31	33	139,65	43	239,03
4	56,84	14	90,85	24	107,13	34	140,23	44	341,99
5	57,60	15	91,61	25	107,34	35	151,79	45	367,54
6	61,48	16	93,11	26	111,47	36	156,57	46	470,25
7	62,80	17	93,46	27	118,67	37	161,89	47	530,66
8	67,67	18	99,01	28	119,81	38	168,91	48	552,66
9	72,80	19	99,63	29	124,03	39	178,67	49	593,90
10	73,55	20	100,96	30	125,67	40	192,92	50	908,20

Tabelle A2.54: Datensatz mit $n = 50$ für Seriensystem Typ $II_{p,0,1}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	36,64	11	89,60	21	113,24	31	145,14	41	219,02
2	38,96	12	94,29	22	113,63	32	149,24	42	226,51
3	66,00	13	94,32	23	114,12	33	169,14	43	238,62
4	68,12	14	96,63	24	114,89	34	175,90	44	250,39
5	71,69	15	102,23	25	118,87	35	188,64	45	265,00
6	74,65	16	104,08	26	119,26	36	189,90	46	280,12
7	78,24	17	104,73	27	122,28	37	192,08	47	303,71
8	78,98	18	106,06	28	138,60	38	192,62	48	343,45
9	80,32	19	107,67	29	139,08	39	192,67	49	526,07
10	83,38	20	109,00	30	143,13	40	210,62	50	722,24

Tabelle A2.55: Datensatz mit $n = 50$ für Seriensystem Typ $II_{p,0,5}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	54,38	11	91,90	21	112,13	31	139,18	41	207,28
2	57,82	12	92,91	22	117,25	32	141,39	42	230,75
3	69,57	13	93,03	23	120,67	33	142,41	43	234,26
4	80,74	14	96,95	24	121,32	34	142,52	44	237,90
5	82,57	15	106,00	25	130,81	35	146,86	45	271,18
6	83,67	16	107,55	26	136,17	36	147,30	46	278,86
7	85,14	17	107,73	27	136,80	37	152,59	47	294,73
8	86,12	18	109,46	28	137,37	38	163,50	48	315,29
9	87,79	19	109,97	29	138,29	39	166,45	49	380,57
10	90,52	20	111,65	30	138,36	40	179,73	50	497,39

Tabelle A2.56: Datensatz mit $n = 50$ für Seriensystem Typ $V_{p,0,1}^{10}$

Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe	Rang-zahl	Rang-größe
1	31,31	11	84,97	21	116,68	31	159,34	41	321,62
2	35,21	12	85,09	22	117,22	32	162,00	42	362,88
3	46,57	13	92,60	23	119,68	33	174,59	43	396,63
4	60,52	14	102,74	24	119,81	34	210,00	44	421,44
5	63,04	15	103,35	25	121,56	35	221,11	45	555,63
6	64,64	16	106,99	26	122,46	36	231,86	46	596,55
7	68,38	17	107,85	27	135,24	37	234,34	47	648,51
8	72,04	18	112,38	28	149,40	38	238,49	48	941,19
9	75,04	19	112,94	29	149,81	39	238,61	49	1577,8
10	76,27	20	116,04	30	153,34	40	272,76	50	1724,6

Tabelle A2.57: Datensatz mit $n = 50$ für Seriensystem Typ $V_p 0,5^{10}$

Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe	Rang- zahl	Rang- größe
1	44,17	11	88,37	21	108,06	31	150,77	41	282,89
2	52,19	12	89,08	22	110,39	32	156,61	42	380,01
3	53,02	13	92,06	23	111,94	33	174,06	43	424,29
4	63,59	14	94,62	24	115,82	34	183,81	44	428,25
5	68,80	15	95,30	25	120,12	35	188,19	45	453,23
6	70,23	16	98,72	26	124,41	36	189,52	46	466,97
7	78,10	17	98,99	27	140,84	37	213,10	47	568,81
8	83,31	18	101,58	28	146,23	38	216,88	48	645,10
9	86,15	19	103,55	29	147,78	39	231,29	49	1470,2
10	88,22	20	104,14	30	150,69	40	262,51	50	1662,6

Lebenslauf

Name Benjamin Klein

geboren am 15.10.1980 in Bietigheim-Bissingen

Schulbildung

1987 – 1991 Hillerschule, Bietigheim-Bissingen

1991 – 1997 Realschule im Aurain, Bietigheim-Bissingen

1997 – 2000 Berufliches Schulzentrum Bietigheim-Bissingen
(Technisches Gymnasium)

Studium

10/01 – 09/07 Maschinenwesen an der Universität Stuttgart
Hauptfächer:
- Konstruktionstechnik
- Landmaschinen

Beruf

10/07 – 03/12 Akademischer Mitarbeiter

Institut für Maschinenelemente der Universität Stuttgart

Seit 05/12 Entwicklungsingenieur bei der Robert Bosch GmbH im Bereich Starter und Generatoren

Liste der bisher erschienenen Berichte aus dem IMA:

Nr.	Verfasser	Titel
1	H.K. Müller	Beitrag zur Berechnung und Konstruktion von Hochdruckdichtungen an schnellaufenden Wellen
2	W. Passera	Konzentrisch laufende Gewinde-Wellen-Dichtung im laminaren Bereich
3	K. Karow	Konzentrische Doppelgewindewellendichtung im laminaren Bereich
3	F.E. Breit	Die Kreiszyinderschalendichtung: Eine Axialspaltdichtung mit druckabhängiger Spaltweite
	W. Sommer	Dichtungen an Mehrphasensystemen: Berührungsfreie Wellendichtungen mit hochviskosen Sperrflüssigkeiten
4	K. Heitel	Beitrag zur Berechnung und Konstruktion konzentrisch und exzentrisch betriebener Gewindewellendichtungen im laminaren Bereich
5	K.-H. Hirschmann	Beitrag zur Berechnung der Geometrie von Evolventenverzahnungen
6	H. Däuble	Durchfluß und Druckverlauf im radial durchströmten Dichtspalt bei pulsierendem Druck
7	J. Rybak	Einheitliche Berechnung von Schneidrädern für Außen- und Innenverzahnungen. Beitrag zu Eingriffsstörungen beim Hohlrad-Verzahnern mittels Schneidräder
8	D. Franz	Rechnergestütztes Entwerfen von Varianten auf der Grundlage gesammelter Erfahrungswerte
9	E. Lauster	Untersuchungen und Berechnungen zum Wärmehaushalt mechanischer Schaltgetriebe
10		Festschrift zum 70. Geburtstag von Prof. Dr.-Ing. K. Talke
11	G. Ott	Untersuchungen zum dynamischen Leckage- und Reibverhalten von Radialwellendichtringen
12	E. Fuchs	Untersuchung des elastohydrodynamischen Verhaltens von berührungsfreien Hochdruckdichtungen
13	G. Sedlak	Rechnerunterstütztes Aufnehmen und Auswerten spannungsoptischer Bilder
14	W. Wolf	Programmsystem zur Analyse und Optimierung von Fahrzeuggetrieben
15	H. v. Eiff	Einfluß der Verzahnungsgeometrie auf die Zahnfußbeanspruchung innen- und außenverzahnter Geradstirnräder
16	N. Messner	Untersuchung von Hydraulikstangendichtungen aus Polytetrafluoräthylen
17	V. Schade	Entwicklung eines Verfahrens zur Einflanken-Wälzprüfung und einer rechnergestützten Auswertemethode für Stirnräder
18	A. Gührer	Beitrag zur Optimierung von Antriebssträngen bei Fahrzeugen
19	R. Nill	Das Schwingungsverhalten loser Bauteile in Fahrzeuggetrieben
20	M. Kammüller	Zum Abdichtverhalten von Radial-Wellendichtringen
21	H. Truong	Strukturorientiertes Modellieren, Optimieren und Identifizieren von Mehrkörpersystemen
22	H. Liu	Rechnergestützte Bilderfassung, -verarbeitung und -auswertung in der Spannungsoptik
23	W. Haas	Berührungsfreie Wellendichtungen für flüssigkeitsbespritzte Dichtstellen
24	M. Plank	Das Betriebsverhalten von Wälzlagern im Drehzahlbereich bis 100.000/min bei Kleinstmengenschmierung
25	A. Wolf	Untersuchungen zum Abdichtverhalten von druckbelastbaren Elastomer- und PTFE-Wellendichtungen
26	P. Waidner	Vorgänge im Dichtspalt wasserabdichtender Gleitringdichtungen
27	Hirschmann u.a.	Veröffentlichungen aus Anlaß des 75. Geburtstags von Prof. Dr.-Ing. Kurt Talke
28	B. Bertsche	Zur Berechnung der Systemzuverlässigkeit von Maschinenbau-Produkten
29	G. Lechner;	Forschungsarbeiten zur Zuverlässigkeit im Maschinenbau
	K.-H.Hirschmann;	
	B. Bertsche	
30	H.-J. Prokop	Zum Abdicht- und Reibungsverhalten von Hydraulikstangendichtungen aus Polytetrafluoräthylen
31	K. Kleinbach	Qualitätsbeurteilung von Kegelradsätzen durch integrierte Prüfung von Tragbild, Einflankenwälzabweichung und Spielverlauf
32	E. Zürn	Beitrag zur Erhöhung der Meßgenauigkeit und -geschwindigkeit eines Mehrkoordinatentasters
33	F. Jauch	Optimierung des Antriebsstranges von Kraftfahrzeugen durch Fahrsimulation
34	J. Grabscheid	Entwicklung einer Kegelrad-Laufprüfmaschine mit thermografischer Tragbilderfassung
35	A. Hölderlin	Verknüpfung von rechnerunterstützter Konstruktion und Koordinatenmeßtechnik
36	J. Kurfess	Abdichten von Flüssigkeiten mit Magnetflüssigkeitsdichtungen
37	G. Borenius	Zur rechnerischen Schädigungsakkumulation in der Erprobung von Kraftfahrzeugteilen bei stochastischer Belastung mit variabler Mittellast
38	E. Fritz	Abdichtung von Maschinenspindeln
39	E. Fritz; W. Haas;	Berührungsfreie Spindelabdichtungen im Werkzeugmaschinenbau. Konstruktionskatalog
	H.K. Müller	

Nr.	Verfasser	Titel
40	B. Jenisch	Abdichten mit Radial-Wellendichtringen aus Elastomer und Polytetrafluorethylen
41	G. Weidner	Klappern und Rasseln von Fahrzeuggetrieben
42	A. Herzog	Erweiterung des Datenmodells eines 2D CAD-Systems zur Programmierung von Mehrkoordinatenmeßgeräten
43	T. Roser	Wissensbasiertes Konstruieren am Beispiel von Getrieben
44	P. Wäschle	Entlastete Wellendichtringe
45	Z. Wu	Vergleich und Entwicklung von Methoden zur Zuverlässigkeitsanalyse von Systemen
46	W. Richter	Nichtwiederholbarer Schlag von Wälzlageeinheiten für Festplattenlaufwerke
47	R. Durst	Rechnerunterstützte Nutprofilentwicklung und clusteranalytische Methoden zur Optimierung von Gewindewerkzeugen
48	G.S. Müller	Das Abdichtverhalten von Gleitringdichtungen aus Siliziumkarbid
49	W.-E. Krieg	Untersuchungen an Gehäuseabdichtungen von hochbelasteten Getrieben
50	J. Grill	Zur Krümmungstheorie von Hüllflächen und ihrer Anwendung bei Werkzeugen und Verzahnungen
51	M. Jäckle	Entlüftung von Getrieben
52	M. Köchling	Beitrag zur Auslegung von geradzahnten Stirnrädern mit beliebiger Flankenform
53	M. Hildebrandt	Schadensfrüherkennung an Wälzkontakten mit Körperschall-Referenzsignalen
54	H. Kaiser	Konstruieren im Verbund von Expertensystem, CAD-System, Datenbank und Wiederholteil-suchsystem
55	N. Stanger	Berührungsfrei abdichten bei kleinem Bauraum
56	R. Lenk	Zuverlässigkeitsanalyse von komplexen Systemen am Beispiel PKW-Automatikgetriebe
57	H. Naunheimer	Beitrag zur Entwicklung von Stufenlosgetrieben mittels Fahrsimulation
58	G. Neumann	Thermografische Tragbilderfassung an rotierenden Zahnrädern
59	G. Wüstenhagen	Beitrag zur Optimierung des Entlasteten Wellendichtrings
60	P. Brodbeck	Experimentelle und theoretische Untersuchungen zur Bauteilzuverlässigkeit und zur Systemberechnung nach dem Booleschen Modell
61	Ch. Hoffmann	Untersuchungen an PTFE-Wellendichtungen
62	V. Hettich	Identifikation und Modellierung des Materialverhaltens dynamisch beanspruchter Flächen-dichtungen
63	K. Riedl	Pulsationsoptimierte Außenzahnpumpen mit ungleichförmig übersetzenden Radpaaren
64	D. Schwuchow	Sonderverzahnungen für Zahnpumpen mit minimaler Volumenstrompulsation
65	T. Spörl	Modulares Fahrsimulationsprogramm für beliebig aufgebaute Fahrzeugtriebstränge und Anwendung auf Hybridantriebe
66	K. Zhao	Entwicklung eines räumlichen Toleranzmodells zur Optimierung der Produktqualität
67	K. Heusel	Qualitätssteigerung von Planetengetrieben durch Selektive Montage
68	T. Wagner	Entwicklung eines Qualitätssysteminformationssystems für die Konstruktion
69	H. Zelßmann	Optimierung des Betriebsverhaltens von Getriebeentlüftungen
70	E. Bock	Schwimmende Wellendichtringe
71	S. Ring	Anwendung der Verzahnungstheorie auf die Modellierung und Simulation des Werkzeug-schleifens
72	M. Klöpfer	Dynamisch beanspruchte Dichtverbindungen von Getriebegehäusen
73	C.-H. Lang	Losteilgeräusche von Fahrzeuggetrieben
74	W. Haas	Berührungsfreies Abdichten im Maschinenbau unter besonderer Berücksichtigung der Fang-labyrinth
75	P. Schiberna	Geschwindigkeitsvorgabe für Fahrsimulationen mittels Verkehrssimulation
76	W. Elser	Beitrag zur Optimierung von Wälzgetrieben
77	P. Marx	Durchgängige, bauteilübergreifende Auslegung von Maschinenelementen mit unscharfen Vorgaben
78	J. Kopsch	Unterstützung der Konstruktionstätigkeiten mit einem Aktiven Semantischen Netz
79	J. Rach	Beitrag zur Minimierung von Klapper- und Rasselgeräuschen von Fahrzeuggetrieben
80	U. Häussler	Generalisierte Berechnung räumlicher Verzahnungen und ihre Anwendung auf Wälzfräserherstellung und Wälzfräsen
81	M. Hüsges	Steigerung der Tolerierungsfähigkeit unter fertigungstechnischen Gesichtspunkten
82	X. Nastos	Ein räumliches Toleranzbewertungssystem für die Konstruktion
83	A. Seifried	Eine neue Methode zur Berechnung von Rollenlagern über lagerinterne Kontakt-Beanspruchungen
84	Ch. Dörr	Ermittlung von Getriebebelastkollektiven mittels Winkelbeschleunigungen
85	A. Veil	Integration der Berechnung von Systemzuverlässigkeiten in den CAD-Konstruktionsprozeß
86	U. Frenzel	Rückenstrukturierte Hydraulikstangendichtungen aus Polyurethan
87	U. Braun	Optimierung von Außenzahnpumpen mit pulsationsarmer Sonderverzahnung
88	M. Lambert	Abdichtung von Werkzeugmaschinen-Flachführungen
89	R. Kubalczyk	Gehäusegestaltung von Fahrzeuggetrieben im Abdichtbereich

Nr.	Verfasser	Titel
90	M. Oberle	Spielbeeinflussende Toleranzparameter bei Planetengetrieben
91	S. N. Dogan	Zur Minimierung der Losteilgeräusche von Fahrzeuggetrieben
92	M. Bast	Beitrag zur werkstückorientierten Konstruktion von Zerspanwerkzeugen
93	M. Ebenhoch	Eignung von additiv generierten Prototypen zur frühzeitigen Spannungsanalyse im Produktentwicklungsprozess
94	A. Fritz	Berechnung und Monte-Carlo Simulation der Zuverlässigkeit und Verfügbarkeit technischer Systeme
95	O. Schrems	Die Fertigung als Versuchsfeld für die qualitätsgerechte Produktoptimierung
96	M. Jäckle	Untersuchungen zur elastischen Verformung von Fahrzeuggetrieben
97	H. Haiser	PTFE-Compounds im dynamischen Dichtkontakt bei druckbelastbaren Radial-Wellendichtungen
98	M. Rettenmaier	Entwicklung eines Modellierungs-Hilfssystems für Rapid Prototyping gerechte Bauteile
99	M. Przybilla	Methodisches Konstruieren von Leichtbauelementen für hochdynamische Werkzeugmaschinen
100	M. Olbrich	Werkstoffmodelle zur Finiten-Elemente-Analyse von PTFE-Wellendichtungen
101	M. Kunz	Ermittlung des Einflusses fahrzeug-, fahrer- und verkehrsspezifischer Parameter auf die Getriebelastkollektive mittels Fahrsimulation
102	H. Ruppert	CAD-integrierte Zuverlässigkeitsanalyse und -optimierung
103	S. Kilian	Entwicklung hochdynamisch beanspruchter Flächendichtverbindungen
104	A. Flaig	Untersuchung von umweltschonenden Antriebskonzepten für Krafffahrzeuge mittels Simulation
105	B. Luo	Überprüfung und Weiterentwicklung der Zuverlässigkeitsmodelle im Maschinenbau mittels Mono-Bauteil-Systemen
106	L. Schüppenhauer	Erhöhung der Verfügbarkeit von Daten für die Gestaltung und Berechnung der Zuverlässigkeit von Systemen
107	J. Ryborz	Klapper - und Rasselgeräuschverhalten von Pkw- und Nkw- Getrieben
108	M. Würthner	Rotierende Wellen gegen Kühlschmierstoff und Partikel berührungsfrei abdichten
109	C. Gitt	Analyse und Synthese leistungsverzweigter Stufenlosgetriebe
110	A. Krolo	Planung von Zuverlässigkeitstests mit weitreichender Berücksichtigung von Vorkenntnissen
111	G. Schöllhammer	Entwicklung und Untersuchung inverser Wellendichtsysteme
112	K. Fronius	Gehäusegestaltung im Abdichtbereich unter pulsierendem Innendruck
113	A. Weidler	Ermittlung von Raffungsfaktoren für die Getriebeerprobung
114	B. Stiegler	Berührungsfreie Dichtsysteme für Anwendungen im Fahrzeug- und Maschinenbau
115	T. Kunstfeld	Einfluss der Wellenoberfläche auf das Dichtverhalten von Radial-Wellendichtungen
116	M. Janssen	Abstreifer für Werkzeugmaschinenführungen
117	S. Buhl	Wechselbeziehungen im Dichtsystem von Radial-Wellendichtring, Gegenlaufläche und Fluid
118	P. Pozsgai	Realitätsnahe Modellierung und Analyse der operativen Zuverlässigkeitskennwerte technischer Systeme
119	H. Li	Untersuchungen zum realen Bewegungsverhalten von Losteilen in Fahrzeuggetrieben
120	B. Otte	Strukturierung und Bewertung von Eingangsdaten für Zuverlässigkeitsanalysen
121	P. Jäger	Zuverlässigkeitsbewertung mechatronischer Systeme in frühen Entwicklungsphasen
122	T. Hitziger	Übertragbarkeit von Vorkenntnissen bei der Zuverlässigkeitstestplanung
123	M. Delonga	Zuverlässigkeitsmanagementsystem auf Basis von Felddaten
124	M. Maisch	Zuverlässigkeitsorientiertes Erprobungskonzept für Nutzfahrzeuggetriebe unter Berücksichtigung von Betriebsdaten
125	J. Orso	Berührungsfreies Abdichten schnelllaufender Spindeln gegen feine Stäube
126	F. Bauer	PTFE-Manschettendichtungen mit Spiralrille - Analyse, Funktionsweise und Erweiterung der Einsatzgrenzen
127	M. Stockmeier	Entwicklung von Klapper- und rasselgeräuschfreien Fahrzeuggetrieben
128	M. Trost	Gesamtheitliche Anlagenmodellierung und -analyse auf Basis stochastischer Netzverfahren
129	P. Lambeck	Unterstützung der Kreativität von verteilten Konstrukteuren mit einem Aktiven Semantischen Netz
130	K. Pickard	Erweiterte qualitative Zuverlässigkeitsanalyse mit Ausfallprognose von Systemen
131	W. Novak	Geräusch- und Wirkungsgradoptimierung bei Fahrzeuggetrieben durch Festradentkopplung
132	M. Henzler	Radialdichtungen unter hoher Druckbelastung in Drehübertragern von Werkzeugmaschinen
133	B. Rzepka	Konzeption eines aktiven semantischen Zuverlässigkeitsinformationssystems
134	C.G. Pflüger	Abdichtung schnelllaufender Hochdruck-Drehübertrager mittels Rechteckring und hocheffizient strukturierter Gleitfläche
135	G. Baitinger	Multiskalenansatz mit Mikrostrukturanalyse zur Drallbeurteilung von Dichtungsgegenläufigen

Nr.	Verfasser	Titel
136	J. Gäng	Berücksichtigung von Wechselwirkungen bei Zuverlässigkeitsanalysen
137	Ch. Maisch	Berücksichtigung der Ölalterung bei der Lebensdauer- und Zuverlässigkeitsprognose von Getrieben
138	D. Kirschmann	Ermittlung erweiterter Zuverlässigkeitsziele in der Produktentwicklung
139	D. Weber	Numerische Verschleißsimulation auf Basis tribologischer Untersuchungen am Beispiel von PTFE-Manschettendichtungen
140	T. Leopold	Ganzheitliche Datenerfassung für verbesserte Zuverlässigkeitsanalysen
141	St. Jung	Beitrag zum Einfluss der Oberflächencharakteristik von Gegenauflflächen auf das tribologische System Radial-Wellendichtung
142	T. Prill	Beitrag zur Gestaltung von Leichtbau-Getriebegehäusen und deren Abdichtung
143	D. Hofmann	Verknüpfungsmodell zuverlässigkeitsrelevanter Informationen in der Produktentwicklung mechatronischer Systeme
144	M. Wacker	Einfluss von Drehungleichförmigkeiten auf die Zahnradlebensdauer in Fahrzeuggetrieben
145	B. Jakobi	Dichtungsgeräusche am Beispiel von Pkw-Lenkungen - Analyse und Abhilfemaßnahmen
146	S. Kiefer	Bewegungsverhalten von singulären Zahnradstufen mit schaltbaren Koppelungseinrichtungen
147	P. Fietkau	Transiente Kontaktberechnung bei Fahrzeuggetrieben