

Dissertation

Optimale Versuchsplanung für Model-Averaging Schätzer

Kira Alhorn
21. Mai 2019

Gutachter: Prof. Dr. Holger Dette
Prof. Dr. Walter Krämer
Prof. Dr. Christine Müller

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

Vorgelegt der Fakultät Statistik der Technischen Universität Dortmund

Dortmund, 21. Mai 2019

Danksagung

Ich bedanke mich bei Prof. Holger Dette und Kirsten Schorning für die ausgezeichnete Betreuung bei der Anfertigung meiner Dissertation. Ihre Begeisterung für die Materie sowie ihre fachliche Kompetenz haben mich stets inspiriert und motiviert. Weiterhin danke ich Prof. Walter Krämer für den Rückhalt und das Vertrauen über viele Jahre hinweg, er hat mich stets auf meinem Weg bestärkt. Prof. Christine Müller danke ich für die Begutachtung dieser Dissertation und Dr. Uwe Ligges für die Übernahme des Vorsitzes der Prüfungskommission.

Ein besonderer Dank gebührt meiner Familie, die mich immer unterstützt hat, insbesondere meiner Mutter, ohne die es diese Dissertation - in so vielerlei Hinsicht - nicht geben würde. Ohne Euch und meine Freunde, die mich nun seit vielen Jahren begleiten, wäre ich nicht dort, wo ich heute bin.

Abschließend möchte ich meine Arbeitskollegen nicht vergessen: Die Gespräche und das Lachen mit Euch haben mir stets neue Energie gegeben. Insbesondere meine Büronachbarn wurden in der Zeit an der Universität für mich unverzichtbar.

Inhaltsverzeichnis

Mathematische Notation	III
Abbildungsverzeichnis	V
Tabellenverzeichnis	VIII
1 Einleitung	1
2 Grundlagen	5
2.1 Grundlegende Konzepte der optimalen Versuchsplanung	5
2.2 Quasi-Maximum-Likelihood Schätzung	10
2.3 Modellsektion und Model-Averaging	14
3 Model-Averaging unter lokalen Alternativen	19
3.1 Das Konzept der lokalen Alternativen	19
3.2 Versuchsplanungskriterien	26
3.3 Anwendung und Simulation	29
3.3.1 Schätzung der ED_α im sigmoidalen Emax-Modell	30
3.3.2 Schätzung der AUC im logistischen Modell	36
3.4 Ein adaptiver Ansatz	41
3.4.1 Schätzung der ED_α im sigmoidalen Emax-Modell	42
3.4.2 Schätzung der AUC im logistischen Modell	44
3.5 Beweise	46
3.6 Weitere Simulationsergebnisse	51
3.6.1 Schätzung der $ED_{0.6}$ im sigmoidalen Emax-Modell	51
3.6.2 Schätzung der AUC im logistischen Modell	55

4	Model-Averaging für nicht-verschachtelte Modelle	59
4.1	Model-Averaging versus Modellselektion	60
4.1.1	Modelle von ähnlicher Form	63
4.1.2	Modelle von unterschiedlicher Form	65
4.2	Asymptotische Verteilung und optimale Versuchsplanung	66
4.2.1	Asymptotische Verteilung	67
4.2.2	Optimale Versuchsplanung	68
4.3	Anwendung und Simulation	72
4.3.1	Modelle von ähnlicher Form	73
4.3.2	Modelle von unterschiedlicher Form	76
4.4	Ein adaptiver Ansatz	78
4.5	Approximation des Versuchsplanungskriteriums	82
4.6	Beweise	86
5	Zusammenfassung und Ausblick	89
	Literaturverzeichnis	91

Mathematische Notation

$AIC(f_s y)$	Wert des AIC-Kriteriums, s. (2.20)
$A_s(\theta_s, \xi), A_s(\theta_s)$	Komponente der Kovarianzmatrix der asymptotischen Vert. des Quasi-Maximum-Likelihood Schätzers, s. (2.17)
$B_{st}(\theta_s, \theta_t, \xi), B_{st}(\theta_s, \theta_t)$	Komponente der Kovarianzmatrix der asymptotischen Vert. des Quasi-Maximum-Likelihood Schätzers, s. (2.18)
δ	Abweichung vom wahren Parametervektor γ , $\delta \in \mathbb{R}^q$
$D\Phi(\xi)(\tilde{\xi} - \xi)$	Richtungsableitung von Φ in ξ in Richtung $\tilde{\xi} - \xi$, s. Def. 2.4
ε_{ij}	Zufällige Messfehler ($i = 1, \dots, k, j = 1, \dots, n_i$)
$\eta(x_i, \vartheta)$	Zur Verteilung von Y_{ij} gehörige Regressionsfunktion, s. (2.1)
$\eta_s(x_i, \vartheta_s)$	Zum Kandidatenmodell f_s gehörige Regressionsfunktion ($s = 1, \dots, r$)
f_s	Dichte des Kandidatenmodells s ($s = 1, \dots, r$)
$g(\cdot x_i)$	Dichte der Verteilung von Y_{ij}
\mathcal{G}	Menge von Dichten g
γ	Zusätzlicher Parametervektor der Dichte g , $\gamma \in \Gamma \subset \mathbb{R}^q$
γ_S	Komponenten von γ , deren Index in S enthalten ist
γ_0	Wahrer Wert von γ
γ_{0,S^c}	Komponenten von γ_0 , deren Index nicht in S enthalten ist
k	Anzahl Trägerpunkte eines Versuchsplans
$KL(g : f_s \theta_s, \xi)$	Kullback-Leibler-Abstand, s. Definition 2.9
$\ell_{n,s}$	Quasi-log-likelihood im Modell f_s ($s = 1, \dots, r$), s. (2.10)
μ	Zielparameter
μ_{true}	Wert des Zielparameters im wahren Modell g
μ_s	Zielparameter im Modell f_s , $\mu_s : \Theta_s \rightarrow \mathbb{R}$ ($s = 1, \dots, r$)
$\hat{\mu}_s$	Schätzer für μ im Modell f_s ($s = 1, \dots, r$), s. (2.12)
$\hat{\mu}_{\text{mav}}$	Model-Averaging Schätzer des Zielparameters, s. (2.22)

n_i	Anzahl unter Versuchsbedingung x_i durchgeführter Versuche ($i = 1, \dots, k$)
n	Anzahl durchgeführter Versuche, $\sum_{i=1}^k n_i = n$
p	Dimension des unbekanntem Parametervektors θ
p_s	Dimension des unbekanntem Parametervektors θ_s
π	A-priori-Verteilung
π_S	Projektionsmatrix, siehe Abschnitt 3.1
q	Dimension des unbekanntem Parametervektors γ
Φ	Versuchsplanungskriterium
r	Anzahl verschiedener Kandidatenmodelle
S	Menge von Indizes, die Komponenten aus γ kennzeichnet, die zu schätzen sind, $S \subset \{1, \dots, q\}$
\mathcal{S}	Menge von Kandidatenmodellen
σ^2	Varianz des Messfehlers ϵ_{ij}
$\sigma_w^2(\theta^*(\xi))$	Varianz des Model-Av. Schätzers bei nicht-verschachtelten Modellen, s. (4.6)
θ	Unbekannter Parametervektor der Dichte g , $\theta \in \Theta \subset \mathbb{R}^p$
θ_0	Wahrer Wert von θ
θ_s	Parametervektor der Dichte f_s ($s = 1, \dots, r$), $\theta_s = (\theta_{s,1}, \dots, \theta_{s,p_s})^T \in \Theta_s \subset \mathbb{R}^{p_s}$
$\hat{\theta}_{n,s}$	(Quasi-)Maximum-Likelihood Schätzer des Par. θ_s ($s = 1, \dots, r$), s. (2.11)
$\theta_{s,g}^*(\xi), \theta_s^*(\xi)$	Parameterwert, der den Kullback-Leibler-Abstand minimiert, s. (2.14)
$\theta^*(\xi)$	Vektor mit $\theta^*(\xi) = (\theta_1^*(\xi)^T, \dots, \theta_r^*(\xi)^T)^T$
ϑ	Unbekannter Parametervektor der Regressionsfunktion η
ϑ_s	Unbekannter Parametervektor der Regressionsfunktion η_s
w_s	Model-Averaging Gew. des Kandidatenmodells f_s ($s = 1, \dots, r$)
$w_s^{\text{AIC}}(Y)$	Model-Averaging Gew. des Schätzers nach Modellselektion ($s = 1, \dots, r$), s. (2.23)
$w_s^{\text{smAIC}}(Y)$	Glatte AIC-Gewichtung, s. (2.24)
\mathcal{X}	Versuchsbereich
ξ	Approximativer Versuchsplan, s. Definition 2.1
ξ_x	Einpunktdesign in x , $x \in \mathcal{X}$, s. (2.6)
Ξ	Menge aller approximativen Versuchspläne
x_i	Trägerpunkt eines Versuchsplans ($i = 1, \dots, k$)
ξ_i	Gewicht des Trägerpunktes x_i in einem Versuchsplan ($i = 1, \dots, k$)
Y_{ij}	Versuchsergebnis unter Versuchsbedingung x_i ($i = 1, \dots, k$) in der j -ten Versuchswiederholung ($j = 1, \dots, n_i$) (Zufallsvariable)
y_{ij}	Realisation von Y_{ij}
Y_n	Vektor der Versuchsergebnisse, $Y_n = (Y_{11}, \dots, Y_{1n_1}, Y_{21}, \dots, Y_{kn_k})^T$
y_n	Realisation von Y_n

Abbildungsverzeichnis

3.1	Erwartungswertfunktion η_4 des sigmoidalen Emax-Modells aus (3.3) für $\vartheta_1 = 1.81$, $\vartheta_2 = 0.79$, $\gamma_1 = 0$ und $\gamma_2 \in \{1, 2, 3\}$	22
3.2	Die Funktion d_π aus (3.22) ausgewertet für den Versuchsplan ξ_A^* aus (3.26) (linke Grafik) und den Versuchsplan ξ_B^* aus (3.30) (rechte Grafik).	31
4.1	Quadratisches und exponentielles Modell mit ihren besten Approximationen durch das log-lineare und das Emax-Modell.	62
4.2	Mittlerer quadratischer Fehler des adaptiven Ansatzes für verschiedene Aufteilungen der Stichprobe. Links: Model-Averaging Schätzer mit glatten AIC-Gewichten. Rechts: Schätzer nach Modellselektion mit dem AIC.	80
4.3	Die Funktion d_π aus (4.19) ausgewertet für den Versuchsplan $\tilde{\xi}_{S_2}$ aus (4.20).	85

Tabellenverzeichnis

3.1	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.6}$ im sigmoidalen Emax-Modell für eine Parameterkombination mit gleichen Modellgewichten.	34
3.2	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.6}$ im sigmoidalen Emax-Modell für eine Parameterkombination mit unterschiedlichen Modellgewichten.	35
3.3	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der AUC im logistischen Modell für eine Parameterkombination mit gleichen Modellgewichten.	38
3.4	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der AUC im logistischen Modell für eine Parameterkombination mit unterschiedlichen Modellgewichten.	40
3.5	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.6}$ im sigmoidalen Emax-Modell für verschiedene Parameterkombinationen mit dem adaptiven Vorgehen.	43
3.6	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der AUC im logistischen Modell für verschiedene Parameterkombinationen mit dem adaptiven Vorgehen.	45
3.7	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.6}$ im sigmoidalen Emax-Modell für verschiedene Parameterkombinationen mit gleichen Modellgewichten.	53
3.8	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.6}$ im sigmoidalen Emax-Modell für verschiedene Parameterkombinationen mit unterschiedlichen Modellgewichten.	54
3.9	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der AUC im logistischen Modell für verschiedene Parameterkombinationen mit gleichen Modellgewichten.	57

3.10	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der AUC im logistischen Modell für verschiedene Parameterkombinationen mit unterschiedlichen Modellgewichten.	58
4.1	Verwendete Modelle und Parameter zum Vergleich von Model-Averaging und Modellselektion bei nicht-verschachtelten Modellen.	61
4.2	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ für eine Menge von Kandidatenmodellen mit ähnlicher Form.	64
4.3	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ für eine Menge von Kandidatenmodellen mit unterschiedlicher Form.	66
4.4	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ für eine Menge von Kandidatenmodellen mit ähnlicher Form und verschiedene Versuchspläne.	74
4.5	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ für eine Menge von Kandidatenmodellen mit unterschiedlicher Form und verschiedene Versuchspläne.	77
4.6	Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ unter Verwendung des adaptiven Vorgehens.	81
4.7	Simulationsergebnis: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0.4}$ unter Verwendung eines approximativ optimalen Versuchsplans.	86

Kapitel 1

Einleitung

In der Medikamentenentwicklung ist die Untersuchung von Dosis-Wirkungs-Beziehungen von zentraler Bedeutung. Gerade in den ersten Phasen von klinischen Studien ist es notwendig, diese möglichst genau zu beschreiben, um so die erforderliche Dosierung eines Medikamentes zu bestimmen, mit der ein gewünschter Effekt erreicht wird. Hierzu werden in verschiedenen Phasen die Medikamente an Probanden getestet. Dabei ist die erste Phase die Untersuchung von pharmakokinetischen- und dynamischen Eigenschaften der Substanzen an wenigen meist gesunden, freiwilligen Probanden. Die zweite Phase dient dazu, erste Erkenntnisse über die Wirksamkeit bei Patienten zu gewinnen. Mittels der ersten Phasen der klinischen Studie kann eine passende Dosierung des Medikamentes für anschließende Phase III und Phase IV Studien ermittelt werden, in denen dann die Wirksamkeit einer neuen Behandlung an einer größeren Patientengruppe nachgewiesen wird (siehe beispielsweise Schumacher und Schulgen (2007)). Dabei ist es nicht nur aus finanziellen, sondern auch aus ethischen Gründen erstrebenswert in allen Phasen der klinischen Studie die Patientenzahlen möglichst klein zu halten.

Ein klassisches Instrument zur Minimierung solcher Fallzahlen ist die optimale Versuchsplanung. Diese blickt auf eine lange Geschichte zurück, wobei exemplarisch die Monographien von Pukelsheim und Rieder (1992), Atkinson et al. (2007) und Fedorov und Leonov (2013) genannt seien, die die Grundlagen der optimalen Versuchsplanung erläutern und einen Überblick über verschiedene Verfahren geben. Unter der Annahme, dass die Regressionsfunktion, die die Dosis-Wirkungs-Beziehung beschreibt, bekannt ist, existieren zahlreiche Ansätze zur optimalen Versuchsplanung. Je präziser das Vorwissen über die Regressionsfunktion und ihre Parameter ist, desto effizienter ist ein basierend hierauf entwickelter Versuchsplan. Dette et al. (2008) zeigen, dass die Effizienz eines Versuchsplans durch eine Missspezifikation des Modells

stark sinken kann. Liegt also kein Vorwissen über das wahre Modell, beispielsweise aus vorangegangenen Phasen einer klinischen Studie, vor, so kann auch ein sorgfältig bestimmter Versuchsplan auf Grundlage einer falschen Modellannahme ungeeignet sein. Ein Ansatz Modellunsicherheit in der Versuchsplanung zu berücksichtigen, geht auf Läuter (1974) zurück. Dieser verwendet für konkurrierende Modelle zur Beschreibung des Einflusses der Versuchsbedingungen auf das Versuchsergebnis ein gemischtes Versuchsplanungskriterium, indem er das Produkt einzelner Kriterien nutzt und so effiziente Versuchspläne für alle betrachteten Modelle bestimmt.

In jeder statistischen Analyse ist ein entscheidender Schritt die Wahl eines passenden Modells, welcher in dem Kriterium aus Läuter (1974) keine Berücksichtigung findet. Hierzu finden sich jedoch zahlreiche weitere Veröffentlichungen im Bereich der optimalen Versuchsplanung. Einen der ersten Ansätze zur optimalen Versuchsplanung zur Unterscheidung verschiedener Modelle bieten Atkinson und Fedorov (1975), indem sie die Trennschärfe eines Tests zur Unterscheidung zweier Kandidatenmodelle mittels der Wahl eines geeigneten Versuchsplans maximieren. Zu den zahlreichen Erweiterungen und weiteren Ansätzen zu dieser Thematik gehören beispielsweise Uciński und Bogacka (2005), López-Fidalgo et al. (2007), Wiens (2009) und Dette und Titoff (2009). Um gleichzeitig eine möglichst gute Trennung zwischen Modellen und eine möglichst präzise Schätzung der enthaltenen Parameter zu ermöglichen, verwenden Zen und Tsai (2002), Atkinson (2008) oder Tommasi (2009) ähnlich wie Läuter (1974) Mischungen aus Versuchsplanungskriterien zur Trennung von Modellen und zur Parameterschätzung.

Die oben genannten Ansätze zur Versuchsplanung legen unterschiedliche Verfahren zur Modellselektion zugrunde. Für einen Überblick über verschiedene häufig verwendete Kriterien zur Modellselektion, wie etwa Akaike's Informationskriterium (AIC), das Bayesianische Informationskriterium (BIC), Mallows C_p , die generalisierte Kreuzvalidierung oder das Fokussierte Informationskriterium (FIC), verweisen wir exemplarisch auf die Monographien von Burnham und Anderson (2002) oder Claeskens und Hjort (2008). Diese Verfahren sind in der Praxis sehr attraktiv, da sie ein einzelnes Modell liefern, in dem die statistische Analyse durchgeführt wird, wodurch eine Interpretation der Ergebnisse leicht fällt. Die einzelnen Modellwahlstrategien sind jeweils unterschiedlich motiviert und haben verschiedene Eigenschaften, welche auch von den betrachteten Modellen abhängen. Hierzu zählen Kriterien wie Konsistenz, Effizienz und Sparsamkeit (jeweils im Sinne von Claeskens und Hjort (2008, Kapitel 4) verwendet). Es gibt jedoch kein universell einsetzbares bestes Modellwahlverfahren, welches sämtliche wünschenswerten Eigenschaften besitzt.

Zudem führen Modellwahlverfahren zu einem *post selection problem* (siehe etwa Poetscher (1991)). Dieses liegt darin begründet, dass Parameterschätzer, die nach einer Modellselektion bestimmt werden, üblicherweise einer Mischung mehrerer Verteilungen folgen. Ist beispielsweise μ ein Zielparameter, der in einem Regressionsmodell zu schätzen ist und ein einzelnes Modell wird mittels eines Modellwahlkriteriums ausgewählt, so führt die Vernachlässigung der zusätzlichen Unsicherheit, die mit der Modellselektion einhergeht, zu Konfidenzintervallen, welche unter Umständen eine zu kleine Überdeckungswahrscheinlichkeit besitzen (siehe Claeskens und Hjort (2008, Kapitel 7)). Die Herleitung von „korrekten“ Konfidenzintervallen ist kompliziert und vom jeweils betrachteten Modell sowie unbekanntem Parametern abhängig (siehe Leeb und Pötscher (2006) für eine Diskussion dieser Problematik).

Als Alternative zu diesem Ansatz schlagen verschiedene Autoren anstelle der Schätzung des Zielparameters in einem einzelnen Modell vor, diesen in allen Kandidatenmodellen zu schätzen und dann über diese Schätzungen zu glätten. Dadurch findet die zusätzliche Variabilität der Schätzer durch die Modellunsicherheit angemessene Berücksichtigung. Vor allem im Bereich der Bayes-Statistik ist dieser Ansatz unter dem Stichwort *bayesianisches Model-Averaging* weit verbreitet (siehe etwa das Tutorium von Hoeting et al. (1999) als ein Beispiel unter vielen). Wie Hjort und Claeskens (2003) herausstellen, ist eine Schwierigkeit bei diesem Ansatz, wie auch bei anderen Bayes-Verfahren, die Wahl geeigneter A-priori Wahrscheinlichkeiten für die Klasse der Kandidatenmodelle. Daher schlagen sie das sogenannte *frequentistische Model-Averaging* als Alternative vor und entwickeln eine asymptotische Theorie für dieses. Ihre Resultate beschränken sich dabei auf eine Modellklasse, die der Annahme *lokaler Alternativen* genügt, wodurch sie eine geschlossene Form der resultierenden asymptotischen Verteilung herleiten können. Simulationsstudien weisen darauf hin, dass die Model-Averaging Schätzung im Vergleich zur Schätzung nach Modellselektion die Schätzgenauigkeit verbessern kann (siehe etwa Breiman (1996) und Raftery und Zheng (2003)). Im Kontext von Dosis-Wirkungsstudien bestätigen Schorning et al. (2016), Aoki et al. (2017) und Buatois et al. (2018) diese Ergebnisse. Daher scheint es vielversprechend, ein Versuchsplanungskriterium zu entwickeln, das diese Art der Parameterschätzung berücksichtigt.

Mit diesem Ziel führen wir in Kapitel 2 zunächst notwendige Grundlagen ein. Wir erläutern dabei grundlegende Konzepte der optimalen Versuchsplanung und treffen Annahmen, die wir zur Bestimmung der asymptotischen Verteilung von Model-Averaging Schätzern benötigen. Ein erster Ansatz zur optimalen Versuchsplanung für Model-Averaging Schätzer findet sich in Kapitel 3. In diesem nutzen wir die Re-

sultate aus Hjort und Claeskens (2003), um Versuchspläne zu bestimmen, die den asymptotischen mittleren quadratischen Fehler des Model-Averaging Schätzers minimieren. Hjort und Claeskens (2003) setzen dabei voraus, dass die Kandidatenmodelle verschachtelt sind und nutzen sogenannte lokale Alternativen. Dadurch handelt es sich bei dem mittleren quadratischen Fehler um einen handlichen Ausdruck, bei dem Verzerrung und Varianz jeweils mit der Stichprobengröße sinken. Wir zeigen anhand von Simulationen, dass die optimalen Versuchspläne die Schätzgenauigkeit des Model-Averaging Schätzers deutlich verbessern können. Zudem ist in den Simulationen zu beobachten, dass Model-Averaging Schätzer präziser sind als Schätzer nach Modellselektion. Um auch Versuchspläne für Model-Averaging Schätzer in allgemeineren Modellklassen bestimmen zu können, leiten wir in Kapitel 4 zunächst die asymptotische Verteilung von Model-Averaging Schätzern unter sehr allgemeinen Annahmen her. Dabei treffen wir nur wenige Restriktionen bezüglich des wahren zugrunde liegenden Modells. In diesem allgemeinen Rahmen zeigt sich nun, dass die Verzerrung des Schätzers unabhängig von der Stichprobengröße ist, während die Varianz mit dieser sinkt. Dadurch erhält für große Stichprobenumfänge die Verzerrung einen größeren Einfluss auf den mittleren quadratischen Fehler als die Varianz und Schätzer nach Modellselektion können präziser sein als Model-Averaging Schätzer. Wir bestimmen optimale Versuchspläne, die den asymptotischen mittleren quadratischen Fehler minimieren und illustrieren auch hier anhand von Simulationsstudien, dass sowohl die Schätzgenauigkeit von Model-Averaging Schätzern als auch die Präzision von Schätzern nach Modellselektion mit diesen Versuchsplänen deutlich erhöht werden kann.

Kapitel 2

Grundlagen

Dieses Kapitel erläutert grundlegende Resultate der statistischen Versuchsplanung sowie wichtige Begriffe im Zusammenhang mit Model-Averaging- und Modellselektionsverfahren, welche zum Verständnis der Arbeit essenziell sind.

2.1 Grundlegende Konzepte der optimalen Versuchsplanung

Wir gehen bei den in dieser Arbeit betrachteten Versuchsplanungsproblemen davon aus, dass insgesamt n Experimente unter k verschiedenen Versuchsbedingungen x_1, \dots, x_k durchgeführt werden können. Unter jeder Versuchsbedingung erheben wir n_i Realisationen y_{i1}, \dots, y_{in_i} ($i = 1, \dots, k$) der Zufallsvariablen Y_{i1}, \dots, Y_{in_i} ($i = 1, \dots, k$), sodass gilt $n = \sum_{i=1}^k n_i$. Die Zufallsvariablen, so wie ihre Realisationen fassen wir jeweils in den Vektoren $Y_n = (Y_{11}, \dots, Y_{1n_1}, Y_{21}, \dots, Y_{kn_k})^T$ und $y_n = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{kn_k})^T$ zusammen. Wir treffen an die wahre Verteilung der Zufallsvariablen die folgende Annahme.

- (A1) Die stochastisch unabhängigen (reellwertigen) Zufallsvariablen Y_{ij} , $i = 1, \dots, k$, $j = 1, \dots, n_i$, haben eine gemeinsame Verteilung mit einer (bezüglich eines Maßes ν) messbaren Radon-Nikodým-Dichte $g(\cdot | x_i)$.

In den Anwendungsbeispielen dieser Arbeit gehen wir konkret davon aus, dass die Zufallsvariablen Y_{ij} Regressionsmodellen der Form

$$Y_{ij} = \eta(x_i, \vartheta) + \varepsilon_{ij}, i = 1, \dots, k, j = 1, \dots, n_i, \quad (2.1)$$

genügen. Die Zufallsvariablen ε_{ij} haben dabei den Erwartungswert $\mathbb{E}(\varepsilon_{ij}) = 0$ sowie die Varianz $\text{Var}(\varepsilon_{ij}) = \sigma^2$ und sind stochastisch unabhängig. Die Dichte $g(\cdot | x_i)$ der

wahren Verteilung hängt somit von den unbekanntem Parametern $\theta^T = (\sigma^2, \vartheta^T) \in \Theta \subset \mathbb{R}^p$ ab. Die Regressionsfunktion η soll im Folgenden differenzierbar bezüglich der Parameter ϑ sowie stetig in x sein.

Wir betrachten, wie es in der approximativen optimalen Versuchsplanung üblich ist (Pukelsheim, 1993, S. 26), den Fall $\lim_{n \rightarrow \infty} \frac{n_i}{n} = \xi_i \in (0, 1)$ für $i = 1, \dots, k$, sodass wir folgende Definition erhalten.

Definition 2.1. Ein *approximativer Versuchsplan* ξ ist ein Wahrscheinlichkeitsmaß mit endlichem Träger auf dem Versuchsbereich \mathcal{X} (Kiefer, 1974). Die Menge aller approximativen Versuchspläne auf dem Versuchsbereich ξ bezeichnen wir mit Ξ . Als Notation für einen solchen Versuchsplan ξ mit *Trägerpunkten* x_1, \dots, x_k und Gewichten ξ_1, \dots, ξ_k verwenden wir die Matrix

$$\xi = \begin{Bmatrix} x_1 & \dots & x_k \\ \xi_1 & \dots & \xi_k \end{Bmatrix}.$$

Bemerkung. Die Gewichte ξ_i eines approximativen Versuchsplans sind als relative Häufigkeiten für die Versuchsbedingung x_i anzusehen. Ist jeweils $\xi_i n$ ganzzahlig, so sind $\xi_i n$ Versuche unter der Versuchsbedingung x_i durchzuführen. In der Praxis müssen jedoch für gewöhnlich die Werte $\xi_i n$ so zu ganzen Zahlen n_i gerundet werden, dass insgesamt $\sum_{i=1}^k n_i = n$ Versuche durchgeführt werden. Dazu nutzt diese Arbeit die im Folgenden erklärte Rundungsprozedur von Pukelsheim und Rieder (1992).

Pukelsheim und Rieder (1992) definieren für ihr Vorgehen zunächst einen Multiplikator κ , den sie verwenden, um Startwerte $\kappa \xi_i$ ($i = 1, \dots, k$) für die Implementierung des Versuchsplans zu erhalten. Sie schlagen vor

$$\kappa = n - \frac{k}{2}$$

zu wählen. Auf die Werte $\kappa \xi_i$ wenden sie zunächst eine Rundungsprozedur an, welche sämtliche nicht ganzzahlige Werte $\kappa \xi_i$ auf die nächstgrößere ganze Zahl rundet, während jeder bereits ganzzahlige Wert entweder unverändert bleibt oder ebenso auf die nächstgrößere ganze Zahl erhöht wird:

$$\tilde{n}_i = \begin{cases} k + 1 & \text{falls } \kappa \xi_i \in (k, k + 1), \\ \{k, k + 1\} & \text{falls } \kappa \xi_i = k, \end{cases} \quad \text{für ein } k \in \mathbb{Z}.$$

Gibt es nun eine Kombination von Kandidaten \tilde{n}_i , sodass $\sum_{i=1}^k \tilde{n}_i = n$ ist, so ist diese Kombination als Versuchsplan zu wählen. Ist allerdings $\sum_{i=1}^k \tilde{n}_i < n$, so müssen weitere Beobachtungen hinzugefügt werden:

$$\tilde{n}_j = \tilde{n}_j + 1, \text{ falls } \tilde{n}_j = \arg \min_{1 \leq i \leq k} \frac{n_i/n}{\xi_i}. \quad (2.2)$$

Die Beobachtung wird also an dem Trägerpunkt x_i hinzugefügt, an dem der relative Anteil an Beobachtungen n_i/n am weitesten vom optimalen Anteil ξ_i entfernt ist. Dieses Verfahren führen wir so lange durch, bis $\sum_{i=1}^k \tilde{n}_i = n$ ist, bis also der Stichprobenumfang voll ausgeschöpft wird. Analog gehen wir vor, falls $\sum_{i=1}^k \tilde{n}_i > n$ ist. In diesem Fall muss jedoch eine Beobachtung entfernt werden:

$$\tilde{n}_j = \tilde{n}_j - 1, \text{ falls } \tilde{n}_j = \arg \max_{1 \leq i \leq k} \frac{(n_i - 1)/n}{\xi_i}. \quad (2.3)$$

Hier wird bei dem Versuchspunkt x_i eine Beobachtung entfernt, bei dem der relative Anteil an Beobachtungen nach Entfernung einer Beobachtung am nächsten an dem Gewicht ξ_i liegt. Auch dies führen wir durch, bis $\sum_{i=1}^k \tilde{n}_i = n$ ist. Sowohl in (2.2) als auch (2.3) muss die Wahl von \tilde{n}_j nicht eindeutig sein, in diesem Fall wählen wir zufällig eines der \tilde{n}_j aus jenen aus, die das Minimum beziehungsweise das Maximum erreichen. Pukelsheim und Rieder (1992) bezeichnen dieses Verfahren als *effizientes Runden*.

Ziel der optimalen Versuchsplanung ist, durch eine genaue Planung der Experimente die statistische Unsicherheit zu minimieren, beispielsweise mit dem Ziel möglichst präzise Parameterschätzungen zu erhalten. Daher basieren viele Versuchsplanungskriterien auf einer Minimierung gewisser Funktionen der Kovarianzmatrix der Schätzung eines unbekanntem Parametervektors (siehe beispielsweise Silvey (1980) für eine Auflistung verschiedener Kriterien). Dazu sind für einen approximativen Versuchsplan ξ gemäß Definition 2.1 sowohl die Anzahl k der Trägerpunkte, als auch die Trägerpunkte x_1, \dots, x_k selbst, sowie die Gewichte ξ_1, \dots, ξ_k so zu bestimmen, dass der Versuchsplan in einem noch zu definierenden Sinne „optimal“ ist.

Um den Begriff der Optimalität zu charakterisieren betrachten wir im Folgenden Operatoren

$$\Phi : \Xi \longrightarrow (0, +\infty],$$

welche eine Abbildung von der Menge aller approximativen Versuchspläne auf dem Versuchsraum \mathcal{X} auf die positiven reellen Zahlen sind. Dabei sei das Ziel im Folgenden stets eine Minimierung dieser Funktion. An die Funktion Φ stellen wir in dieser Arbeit keine weiteren Bedingungen, sodass es sich nicht um eine Informationsfunktion im Sinne von Pukelsheim (1993, Kapitel 5) handelt, insbesondere ist die Funktion nicht notwendigerweise konvex.

Beispiel 2.2. Spätestens seit Kiefer und Wolfowitz (1960) ist eines der am häufigsten verwendeten Optimalitätskriterien in der Versuchsplanung die D -Optimalität.

Zahlreiche Standardlehrwerke zur optimalen Versuchsplanung behandeln dieses Kriterium ausführlich, siehe beispielsweise Silvey (1980) oder Fedorov (1972). Trotz seiner langen Geschichte ist dieses Kriterium immer noch Gegenstand vieler Untersuchungen, siehe etwa Atkinson (2008) oder Dette et al. (2010). Dieses Kriterium basiert auf der Kovarianzmatrix $A(\xi, \theta)$ der Parameterschätzungen des unbekannt Parameters $\theta \in \mathbb{R}^p$ der wahren Verteilung mit Dichte g . Ein D -optimaler Versuchsplan minimiert die Determinante dieser Matrix. Die Kovarianzmatrix ist dabei im Allgemeinen von dem unbekannt Parameter θ sowie dem Versuchsplan ξ abhängig.

Definition 2.3. Ein Versuchsplan $\xi^* \in \Xi$ heißt *D-optimal*, falls gilt

$$\Phi_D(\xi^*, \theta) = \min_{\xi \in \Xi} \Phi_D(\xi, \theta)$$

mit

$$\Phi_D(\xi, \theta) = \begin{cases} |A(\xi, \theta)|, & A(\xi, \theta) \text{ nicht singulär,} \\ \infty, & A(\xi, \theta) \text{ singulär.} \end{cases}$$

Dabei bezeichnet $|\cdot|$ die Determinante einer Matrix und $A(\xi, \theta)$ die Kovarianzmatrix eines Schätzers $\hat{\theta}$ für θ .

In den vorliegenden Versuchsplanungsproblemen ist es im Allgemeinen nicht möglich, analytische Lösungen für die Optimierungsprobleme zu finden. Daher ist es hilfreich Optimalitätsbedingungen herzuleiten, welche eine Überprüfung der numerischen Lösung auf Optimalität erlauben. Hierzu kann die Richtungsableitung des Operators Φ genutzt werden.

Definition 2.4. (Papageorgiou und Winkert, 2018, Definition 5.2.1.)

Sei X ein normierter Vektorraum (im Sinne von Papageorgiou und Winkert (2018, Definition 3.1.13.)), $\Xi \subset X$ offen, $\xi \in \Xi$ sowie $\Phi : \Xi \rightarrow \mathbb{R}$. Dann ist für $\tilde{\xi} \in \Xi$ die Richtungsableitung von Φ in ξ in Richtung $\tilde{\xi} - \xi$ gegeben durch

$$D\Phi(\xi)(\tilde{\xi} - \xi) = \lim_{\alpha \downarrow 0} \frac{\Phi(\xi + \alpha(\tilde{\xi} - \xi)) - \Phi(\xi)}{\alpha}, \quad (2.4)$$

sofern der Grenzwert existiert.

Bemerkung. Der normierte Vektorraum X aus Definition 2.4 ist im vorliegenden Fall der Raum der signierten Maße versehen mit der Totalvariationsnorm, von welchem die Menge Ξ aller Wahrscheinlichkeitsmaße mit endlichem Träger auf \mathcal{X} eine Teilmenge bildet (siehe Klenke (2013, Definition 7.40 und Korollar 7.45)).

Mittels der Richtungsableitung (2.4) ist es möglich, eine notwendige Bedingung für das Vorliegen eines Minimums zu konstruieren.

Satz 2.5. *Gegeben sei ein Versuchsplan $\xi^* \in \Xi$ mit $\Phi(\xi^*) \leq \Phi(\xi)$ für alle $\xi \in \Xi$. Dann gilt für die Richtungsableitung von Φ in ξ^* in alle Richtungen $\tilde{\xi} - \xi^*$ mit $\tilde{\xi} \in \Xi$ die Ungleichheit*

$$D\Phi(\xi^*)(\tilde{\xi} - \xi^*) \geq 0, \quad (2.5)$$

sofern die Richtungsableitung existiert.

Beweis. Ist Φ minimal in ξ^* , so gilt

$$\Phi(\xi^* + \alpha(\tilde{\xi} - \xi^*)) - \Phi(\xi^*) \geq 0$$

für alle $\alpha \in [0, 1]$. Damit ist aber auch der Zähler in (2.4) nicht-negativ und (2.5) folgt unmittelbar. \square

Bemerkung. Die Ungleichheit (2.5) gilt für alle Versuchspläne $\tilde{\xi} \in \Xi$, insbesondere also auch für Einpunktdesigns

$$\xi_x = \left\{ \begin{array}{c} x \\ 1 \end{array} \right\}. \quad (2.6)$$

Da sämtliche Versuchspläne $\tilde{\xi} \in \Xi$ über Konvexkombinationen von Einpunktdesigns darstellbar sind, können wir mit

$$D\Phi(\xi^*)(\xi_x - \xi^*) \geq 0 \quad \forall x \in \mathcal{X} \quad (2.7)$$

eine zu (2.5) äquivalente Bedingung für das Vorliegen eines Minimums angeben. Diese Bedingung können wir grafisch überprüfen, sodass die Verwendung von (2.7) attraktiver ist als die Verwendung von (2.5).

Es ist zu beachten, dass obige Aussagen analog gelten, wenn das Kriterium Φ zu maximieren ist. In diesem Fall betrachten wir eine Minimierung von $-\Phi$.

Beispiel 2.6. (Fortsetzung von Beispiel 2.2)

Wir betrachten das Kriterium

$$\Phi_D(\xi) = \begin{cases} \log|A(\theta, \xi)|, & A(\theta, \xi) \text{ nicht singulär,} \\ \infty, & A(\theta, \xi) \text{ singulär,} \end{cases} \quad (2.8)$$

welches eine Transformation des Kriteriums aus Definition 2.3 ist. Für dieses erhalten wir

$$D\Phi_D(\xi^*)(\xi_x - \xi^*) = \text{tr}(A(\theta, \xi^*)^{-1}A(\theta, \xi_x)) - p, \quad (2.9)$$

wobei p die Dimension des unbekanntenen Parameters θ ist und tr die Spur einer Matrix bezeichnet (siehe Fedorov und Leonov (2013, S. 68)).

Klassische Versuchsplanungskriterien sind oftmals konvex. Dabei ist die Funktion Φ konvex, falls für $\xi_1, \xi_2 \in \Xi$ und $\lambda \in [0, 1]$ gilt

$$\Phi(\lambda\xi_1 + (1 - \lambda)\xi_2) \leq \lambda\Phi(\xi_1) + (1 - \lambda)\Phi(\xi_2)$$

(Pshenichnyi, 1971, Definition 12). Konvexe Versuchsplanungskriterien sind insbesondere deswegen beliebt, da für sie die Bedingung aus Satz 2.5 nicht nur notwendig, sondern auch hinreichend ist.

Satz 2.7. *Für eine konvexe und beschränkte Funktion Φ existiert die Richtungsableitung (2.4). Ist weiterhin die Bedingung (2.7) für ein $\xi^* \in \Xi$ erfüllt, so minimiert der Versuchsplan ξ^* das Kriterium Φ .*

Beweis. Pshenichnyi (1971, Theorem 1.2. und Theorem 2.3.). □

Beispiel 2.8. (Fortsetzung von Beispiel 2.2)

Das Kriterium aus Definition 2.3 ist nicht konvex. Die Transformation aus (2.8) hingegen schon (siehe Fedorov und Leonov (2013, S. 61)). Somit ist mit Satz 2.7 und (2.9) ein Versuchsplan $\xi^* \in \Xi$ genau dann D -optimal, falls die Ungleichung

$$\text{tr}(A(\theta, \xi^*)^{-1}A(\theta, \xi_x)) \geq p$$

für alle $x \in \mathcal{X}$ erfüllt ist.

2.2 Quasi-Maximum-Likelihood Schätzung

In Abschnitt 2.1 haben wir stets angenommen, dass die Dichte $g(\cdot | x_i)$ der Verteilung der Zufallsvariablen Y_{ij} bekannt ist. In der Praxis ist die wahre Verteilung sowie ihre Dichte für gewöhnlich jedoch nicht bekannt. Stattdessen liegen Kandidatenmodelle vor, welche das wahre Modell möglichst gut beschreiben sollen. Wir gehen davon aus, dass r verschiedene Kandidatenmodelle vorliegen und treffen für sie die folgende Annahme. Um die Notation zu vereinfachen bezeichne dabei Y die Zufallsvariable, die das Ergebnis eines Versuches beschreibt; und $g(y | x)$ ihre Dichte für eine Versuchsbedingung $x \in \mathcal{X}$.

- (A2) Jedes Kandidatenmodell besitzt eine Verteilung, welche (bezüglich des Maßes ν) eine Radon-Nikodým-Dichte $f_s(\cdot | x, \theta_s)$ mit Parameter $\theta_s \in \Theta_s \subset \mathbb{R}^{p_s}$ besitzt, wobei Θ_s kompakt ist. Diese ist für alle $\theta_s \in \Theta_s$ messbar in $y \in \mathbb{R}$ und für alle $y \in \mathbb{R}$ stetig in $\theta_s \in \Theta_s$. Wir bezeichnen das s -te Kandidatenmodell auch mit f_s .

In den Anwendungsbeispielen dieser Arbeit verwenden wir für die Kandidatenmodelle ebenfalls Regressionsmodelle der Form (2.1), wobei wir jedoch davon ausgehen, dass die wahre Form η der Regressionsfunktion unbekannt ist. Im Kandidatenmodell f_s ersetzen wir diese daher durch eine Regressionsfunktion η_s ($s = 1, \dots, r$), welche die gleichen Annahmen erfüllt wie η und einen Parameter ϑ_s enthält. Die unbekannt Parameter fassen wir zusammen zu $\theta_s^T = (\sigma^2, \vartheta_s^T) \in \Theta_s \subset \mathbb{R}^{p_s}$, sodass die Dichte im Modell s gegeben ist durch $f_s(\cdot | x, \theta_s)$ ($s = 1, \dots, r$). Im Allgemeinen ist dabei der Parameter θ_s unbekannt.

Basierend auf dem Beobachtungsvektor y_n , der gemäß eines entsprechenden Versuchsplans erhoben wurde, können die unbekannt Parameter der Kandidatenmodelle mit der (Quasi-)Maximum-Likelihood Methode geschätzt werden. Dazu definieren wir die *Quasi-log-Likelihood* im Kandidatenmodell f_s als

$$\ell_{n,s}(\theta_s | y_n) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \log f_s(y_{ij} | x_i, \theta_s), \quad (2.10)$$

wobei wir hier die Unabhängigkeit der Zufallsvariablen Y_{11}, \dots, Y_{kn_k} ausgenutzt haben. Der Quasi-Maximum-Likelihood Schätzer ist dann derjenige Parameterwert, welcher (2.10) maximiert, sprich

$$\hat{\theta}_{n,s} = \arg \max_{\theta_s \in \Theta_s} \ell_{n,s}(\theta_s | y_n). \quad (2.11)$$

Wir gebrauchen die Terminologie „quasi“, da wir nicht die tatsächliche Likelihood verwenden. Diese erhielte man unter Verwendung der wahren Dichte $g(\cdot | x)$ (Woolridge, 2010, S. 503). Unter den Annahmen (A1) und (A2) zeigt White (1982, Satz 2.1), dass für alle n ein messbarer Quasi-Maximum-Likelihood Schätzer gemäß (2.11) existiert.

Oftmals ist nicht die Schätzung des Parameters θ_s selbst von Interesse, sondern die Schätzung eines *Zielparameters* μ . Dieser Zielparameter wiederum ist im Kandidatenmodell f_s für gewöhnlich darstellbar über eine Funktion $\mu_s : \Theta_s \rightarrow \mathbb{R}$. Die Funktionen μ_s seien jeweils stetig differenzierbar bezüglich θ_s . Mit der Quasi-Maximum-Likelihood Schätzung aus (2.11) erhalten wir eine Schätzung für den Zielparameter μ im Modell f_s , nämlich

$$\hat{\mu}_s = \mu_s(\hat{\theta}_{n,s}). \quad (2.12)$$

Um die Verteilung solcher Schätzungen zu bestimmen, nutzen wir in den folgenden Kapiteln Resultate aus der (Quasi-)Maximum-Likelihood Theorie, welche wir hier aufführen. Sämtliche Annahmen beziehen sich dabei auf Indizes $s, t \in \{1, \dots, r\}$.

Wir benötigen zunächst den Kullback-Leibler Abstand der Modelle f_s ($s = 1, \dots, r$) zum wahren Modell g .

Definition 2.9. Für einen approximativen Versuchsplan ξ ist der Kullback-Leibler-Abstand vom Kandidatenmodell f_s zum wahren Modell g gegeben durch

$$\text{KL}(g : f_s \mid \theta_s, \xi) = \int \int g(y \mid x) \log \left(\frac{g(y \mid x)}{f_s(y \mid x, \theta_s)} \right) dy d\xi(x). \quad (2.13)$$

Der Kullback-Leibler-Abstand entspricht der (unter g) erwarteten Differenz zwischen den logarithmierten Dichten g und f_s . Um die Existenz des Kullback-Leibler-Abstands sowie die Identifizierbarkeit des Parameters θ_s zu gewährleisten, benötigen wir die folgende Annahme.

(A3) Für alle $x \in \mathcal{X}$ existiert $\mathbb{E}_{g(\cdot|x)}(\log(g(Y|x)))$ und für jedes Kandidatenmodell wird die Funktion $y \mapsto |\log f_s(y \mid x, \theta_s)|$ von einer Funktion dominiert, welche integrierbar bezüglich $g(\cdot \mid x)$ und unabhängig von θ_s ist. Zudem hat der Kullback-Leibler-Abstand (2.13) ein eindeutiges Minimum

$$\theta_{s,g}^*(\xi) = \arg \min_{\theta_s \in \Theta_s} \text{KL}(g : f_s \mid \theta_s, \xi). \quad (2.14)$$

Ist in (2.14) die Abhängigkeit von der zugrundeliegenden wahren Dichte g eindeutig, so schreiben wir auch $\theta_s^*(\xi)$ anstelle von $\theta_{s,g}^*(\xi)$.

Satz 2.10. *Unter den Annahmen (A1)-(A3) ist der Quasi-Maximum-Likelihood Schätzer aus (2.11) stark konsistent für den Wert $\theta_{s,g}^*(\xi)$, das heißt*

$$\hat{\theta}_{n,s} \xrightarrow{\text{a.s.}} \theta_{s,g}^*(\xi),$$

wobei $\xrightarrow{\text{a.s.}}$ fast sichere Konvergenz bezeichnet. Somit konvergiert der Quasi-Maximum-Likelihood Schätzer gegen den Wert, der den Kullback-Leibler-Abstand zum wahren Modell minimiert.

Beweis. Siehe White (1982, Satz 2.2). □

Um die asymptotische Verteilung des Quasi-Maximum-Likelihood Schätzers anzugeben, benötigen wir zusätzliche Annahmen. Dazu sei zunächst die Ableitung der logarithmierten Dichte f_s ($s = 1, \dots, r$) bezüglich des Parameters $\theta_s = (\theta_{s,1}, \dots, \theta_{s,p_s})^T$ bezeichnet mit

$$\frac{\partial \log f_s(y \mid x, \theta_s)}{\partial \theta_s} = \begin{pmatrix} \frac{\partial \log f_s(y|x, \theta_s)}{\partial \theta_{s,1}} \\ \vdots \\ \frac{\partial \log f_s(y|x, \theta_s)}{\partial \theta_{s,p_s}} \end{pmatrix}. \quad (2.15)$$

An diese Ableitung stellen wir die folgende Forderung.

(A4) Für alle $x \in \mathcal{X}$ ist die Funktion $y \mapsto \frac{\partial \log f_s(y|x, \theta_s)}{\partial \theta_s}$ aus (2.15) messbar für alle $\theta_s \in \Theta_s$ und zudem stetig differenzierbar bezüglich θ_s für alle $y \in \mathbb{R}$.

Die Differenzierbarkeit aus (A4) nutzen wir, um basierend auf der Ableitung (2.15) durch erneutes Differenzieren die Hesse-Matrix zu erhalten:

$$\frac{\partial^2 \log f_s(y | x, \theta_s)}{\partial \theta_s \partial \theta_s^T} = \begin{pmatrix} \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,1} \partial \theta_{s,1}} & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,1} \partial \theta_{s,2}} & \cdots & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,1} \partial \theta_{s,p_s}} \\ \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,2} \partial \theta_{s,1}} & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,2} \partial \theta_{s,2}} & \cdots & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,2} \partial \theta_{s,p_s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,p_s} \partial \theta_{s,1}} & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,p_s} \partial \theta_{s,2}} & \cdots & \frac{\partial^2 \log f_s(y|x, \theta_s)}{\partial \theta_{s,p_s} \partial \theta_{s,p_s}} \end{pmatrix} \in \mathbb{R}^{p_s \times p_s}. \quad (2.16)$$

Mit dieser ergibt sich eine weitere Annahme.

(A5) Die Absolutbeträge der Einträge der (matrixwertigen) Funktionen

$$y \mapsto \frac{\partial^2 \log f_s(y | x, \theta_s)}{\partial \theta_s \partial \theta_s^T}$$

aus (2.16) und die Absolutbeträge der Einträge der Matrix

$$y \mapsto \frac{\partial \log f_s(y | x, \theta_s)}{\partial \theta_s} \left(\frac{\partial \log f_t(y | x, \theta_t)}{\partial \theta_t} \right)^T$$

werden für alle $y \in \mathbb{R}$ und $\theta_s \in \Theta_s, \theta_t \in \Theta_t$ von einer Funktion dominiert, welche integrierbar bezüglich $g(\cdot | x)$ ist.

Durch die Annahmen (A4) und (A5) ist die Existenz der folgenden Matrizen für $s, t \in \{1, \dots, r\}$ gesichert:

$$A_s(\theta_s, \xi) = \sum_{i=1}^k \xi_i \mathbb{E}_{g(\cdot | x_i)} \left(\frac{\partial^2 \log f_s(Y_{ij} | x_i, \theta_s)}{\partial \theta_s \partial \theta_s^T} \right), \quad (2.17)$$

$$B_{st}(\theta_s, \theta_t, \xi) = \sum_{i=1}^k \xi_i \mathbb{E}_{g(\cdot | x_i)} \left(\frac{\partial \log f_s(Y_{ij} | x_i, \theta_s)}{\partial \theta_s} \left(\frac{\partial \log f_t(Y_{ij} | x_i, \theta_t)}{\partial \theta_t} \right)^T \right). \quad (2.18)$$

Hierbei verwenden wir die Erwartungswerte bezüglich der wahren Verteilung mit Dichte $g(\cdot | x)$. Ist die Abhängigkeit vom Versuchsplan ξ eindeutig, so verwenden wir als Kurzschreibweise auch $A_s(\theta_s)$ und $B_{st}(\theta_s, \theta_t)$. Für diese Matrizen soll Folgendes gelten.

(A6) Für die Matrizen aus (2.17), (2.18) und den Vektor aus (2.14) treffen wir folgende Annahmen:

- (1) Der Vektor $\theta_s^*(\xi)$ ist ein innerer Punkt der Menge Θ_s ,
- (2) Die Matrix $B_{ss}(\theta_s^*(\xi), \theta_s^*(\xi))$ ist nicht singulär,
- (3) Die Matrix $A_s(\theta_s^*(\xi))$ ist nicht singulär.

Satz 2.11. *Unter den Annahmen (A1)-(A6) ist der Quasi-Maximum-Likelihood Schätzer asymptotisch normalverteilt ist. Es gilt*

$$\sqrt{n}(\hat{\theta}_{n,s} - \theta_s^*(\xi)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A_s^{-1}(\theta_s^*(\xi))B_{ss}(\theta_s^*(\xi), \theta_s^*(\xi))A_s^{-1}(\theta_s^*(\xi))), \quad (2.19)$$

wobei $\xrightarrow{\mathcal{D}}$ die Konvergenz in Verteilung bezeichnet.

Beweis. Siehe White (1982, Satz 3.2). □

Bemerkung. White (1982) verwendet anstelle von (A6)(3) die schwächere Annahme, dass für alle $\bar{\theta}_s$ in einer offenen Umgebung von $\theta_s^*(\xi)$ die Matrix $A_s(\bar{\theta}_s, \xi)$ konstanten Rang hat und zeigt, dass dann unter den übrigen Annahmen $A_s(\theta_s^*(\xi), \xi)$ nicht singulär ist.

Befindet sich unter den Kandidatenmodellen das wahre Modell, gilt also $g(\cdot | x) = f_s(\cdot | x, \theta_s^0)$ für ein $\theta_s^0 \in \Theta_s$, so erhalten wir die klassischen Maximum-Likelihood Resultate. Insbesondere gilt $\theta_s^*(\xi) = \theta_s^0$, sodass der Maximum-Likelihood Schätzer nach Satz 2.10 stark konsistent für den wahren Parameterwert ist. Weiterhin ist unter schwachen Regularitätsbedingungen (siehe (A7) in White (1982)) $A_s(\theta_s^0) = -B_{ss}(\theta_s^0, \theta_s^0)$ und $-A_s(\theta_s^0)$ ist die *Fisher-Informationmatrix* (White, 1982, Theorem 3.3).

2.3 Modellselektion und Model-Averaging

In Abschnitt 2.2 ist das Verhalten des Quasi-Maximum-Likelihood Schätzers $\hat{\theta}_{n,s}$ in einem Kandidatenmodell f_s bei wahren Modell g erläutert. Genauer gesagt liegen in dieser Arbeit r Kandidatenmodelle vor, in denen ein Zielparameter μ mittels (2.12) geschätzt werden kann. Eine klassische Vorgehensweise in dieser Situation ist die Berechnung eines Modellwahlkriteriums und mittels der realisierten Werte des Kriteriums die Bestimmung eines „besten“ Modells f_{s^*} , wodurch eine einzelne Schätzung $\hat{\mu}_{s^*}$ resultiert. Ein populäres Modellwahlkriterium ist Akaikes Informationskriterium (AIC), welches in Standardlehrwerken zur Modellselektion ausführlich behandelt wird (siehe etwa Claeskens und Hjort (2008) und Burnham und Anderson (2002)). Dieses von Akaike (1973) eingeführte Kriterium basiert auf dem Kullback-Leibler-Abstand (2.13) der Modelle f_s ($s = 1, \dots, r$) zum wahren Modell g . Mittels

des AIC-Kriteriums wird der Kullback-Leibler-Abstand der Kandidatenmodelle zum unbekanntem wahren Modell geschätzt und das Modell ausgewählt, welches den geringsten geschätzten Abstand aufweist. Das resultierende Modellwahlkriterium ist nach Claeskens und Hjort (2008, Kapitel 2.1.) für das Modell f_s und Beobachtungen $y_n \in \mathbb{R}^n$ gegeben durch

$$\text{AIC}(f_s|y_n) = 2\ell_{n,s}(\hat{\theta}_{n,s}|y_n) - 2p_s. \quad (2.20)$$

Das Modell, in dem der Wert des Kriteriums maximal ist, wird zur Schätzung des Parameters verwendet. Somit bewirkt die Verwendung dieses Kriteriums eine Balance zwischen einer guten Modellanpassung, welche sich durch einen großen Wert der Quasi-log-Likelihood ausdrückt, und der Modellkomplexität, gemessen anhand der Anzahl p_s der zu schätzenden Parameter.

Ein weiteres populäres Modellwahlkriterium, das dem AIC von der Form ähnlich ist, ist das Bayesianische Informationskriterium (BIC). Dieses Kriterium approximiert die A-posteriori Wahrscheinlichkeit für ein Kandidatenmodell f_s für gegebene Beobachtungen $y_n \in \mathbb{R}^n$ mittels

$$\text{BIC}(f_s|y_n) = 2\ell_{n,s}(\hat{\theta}_{n,s}|y_n) - (\log n)p_s. \quad (2.21)$$

Auch hier wird das Modell s^* ausgewählt, welches den höchsten Wert $\text{BIC}(f_{s^*}|y_n)$ und somit die größte geschätzte A-posteriori Wahrscheinlichkeit innerhalb der Kandidatenmodelle besitzt. Mittels dieses Kriteriums werden weniger komplexe Modelle ausgewählt als mittels des AIC, da hier die Anzahl der Parameter stärker bestraft wird (Claeskens und Hjort, 2008, Kapitel 3).

Insbesondere für das AIC existieren zahlreiche Erweiterungen und Korrekturen und es gibt eine Fülle weiterer Modellwahlkriterien, die ebenso wie AIC und BIC auf informationstheoretischen Ansätzen beruhen (siehe Burnham und Anderson (2002) für einen Überblick über solche). Auch Verfahren, die nicht auf Informationskriterien beruhen, sind in der Modellwahl gebräuchlich, etwa die Verwendung von Tests für die Anpassungsgüte eines Modells. Einen Überblick hierzu bieten Claeskens und Hjort (2008).

An Modellselektionsverfahren können unterschiedliche Anforderungen gestellt werden. Hierzu zählen Konsistenz, Sparsamkeit und Effizienz. Claeskens und Hjort (2008) bezeichnen ein Modellselektionsverfahren als *schwach konsistent*, falls die Wahrscheinlichkeit gegen 1 konvergiert, dass das Verfahren das Modell auswählt, das den geringsten Kullback-Leibler-Abstand zum wahren Modell besitzt. Gibt es jedoch mehrere Modelle mit dem gleichen Kullback-Leibler-Abstand, so ist das *sparsamere* Modell zu bevorzugen, welches weniger Parameter beinhaltet. Claeskens und

Hjort (2008) erläutern, dass sowohl das AIC als auch das BIC schwach konsistent sind, jedoch nur das BIC auch sparsam ist. Neben der Konsistenz erörtern Claeskens und Hjort (2008) ebenso das Konzept der *Effizienz*. Hierzu nutzen sie den mittleren quadratischen Fehler und nennen ein Kriterium effizient, wenn der Quotient des Fehlers im gewählten Modell und des Fehlers im (bezüglich des Kullback-Leibler-Abstands) besten Modell in Wahrscheinlichkeit gegen 1 konvergiert. Die Effizienz eines Kriteriums hängt dabei stark von der betrachteten Modellklasse ab. So ist beispielsweise das AIC effizient im Kontext von klassischen linearen Regressionsmodellen und autoregressiven Modellen. Jedoch kann keines der Modellwahlkriterien sämtliche Anforderungen gleichzeitig erfüllen, sodass es kein universell bestes Modellwahlkriterium gibt. Es ist zudem hinlänglich bekannt, dass die Modellwahl mit anschließender Parameterschätzung zu einer größeren Variabilität der Schätzung führt, welche bei anschließender Inferenz zu berücksichtigen ist (siehe beispielsweise Poetscher (1991) als eine der ersten Veröffentlichungen zu dieser Problematik oder Claeskens und Hjort (2008) für eine ausführliche Abhandlung). Beispielsweise führt die Ignoranz der zusätzlichen Variabilität bei der Konstruktion von Konfidenzintervallen zu Intervallen mit zu kleiner Überdeckungswahrscheinlichkeit. Als Alternative entwickeln Hjort und Claeskens (2003) mit ihren *frequentistischen Model-Averaging Schätzern* Rahmenbedingungen, um die Verteilung solcher Schätzer erklären zu können. Gleichzeitig bieten sie hiermit eine andere Möglichkeit zur Schätzung des Zielparameters. Sie verwenden die Schätzung $\hat{\mu}_s$ aus (2.12) und definieren den Model-Averaging Schätzer $\hat{\mu}_{\text{mav}}$ als

$$\hat{\mu}_{\text{mav}} = \sum_{s=1}^r w_s \hat{\mu}_s, \quad (2.22)$$

wobei w_1, \dots, w_r Gewichte für die einzelnen Modelle darstellen mit $\sum_{s=1}^r w_s = 1$. Hjort und Claeskens (2003) erlauben dabei, dass diese Gewichte von den Daten abhängen. Aufgrund dieser Abhängigkeit nutzen wir auch die Bezeichnung *zufällige Gewichte*. Über eine geeignete Wahl der Gewichte können Hjort und Claeskens (2003) auch Modellwahlverfahren beschreiben. Für das obige Beispiel der Modellwahl mittels des AIC erreichen sie dies, indem sie die Gewichtung

$$w_s^{\text{AIC}}(Y_n) = \mathbb{1}\{f_s = f_{s^*}\} \quad (2.23)$$

verwenden. Dabei stellt $\mathbb{1}\{A\}$ die Indikatorfunktion für die Menge A dar und f_{s^*} ist dasjenige Modell, welches unter den Kandidatenmodellen den größten AIC Wert besitzt. Soll Model-Averaging anstelle von Modellwahl verwendet werden und es liegen keine A-priori Informationen vor, so sind gleiche Modellgewichte eine naheliegende

Wahl. Diese nutzen beispielsweise Stock und Watson (2004) und Kapetanios et al. (2008). Ausgefeiltere Gewichte schlagen unter Anderem Claeskens und Hjort (2008) vor, indem sie über die verschiedenen AIC Werte glätten:

$$w_s^{\text{smAIC}}(Y_n) = \frac{\exp(\frac{1}{2}\text{AIC}(f_s|Y_n))}{\sum_{s=1}^r \exp(\frac{1}{2}\text{AIC}(f_s|Y_n))}. \quad (2.24)$$

Diese Gewichtung nennen wir *glatte AIC-Gewichtung*. Die glatten AIC-Gewichte finden häufig Anwendung in der Praxis, etwa in Verrier et al. (2014) im Kontext von Dosis-Wirkungs-Studien oder in Kapetanios et al. (2008) zur Vorhersage der Inflation in Großbritannien. Analog zu (2.23) und (2.24) können Gewichte für das BIC-Kriterium (2.21) konstruiert werden. Weitere Gewichte, die zur Bestimmung von Model-Averaging Schätzern gebräuchlich sind, finden sich etwa in Buckland et al. (1997), Hjort und Claeskens (2003) oder Hansen (2007). Die Verteilung von Model-Averaging Schätzern der Form (2.22) findet sich unter speziellen Annahmen in Hjort und Claeskens (2003). Für feste Modellgewichte geben wir diese Resultate in Kapitel 3 wieder und erweitern sie in Kapitel 4 auf die in Abschnitt 2.1 und Abschnitt 2.2 getroffenen Annahmen.

Bemerkung. Claeskens und Hjort (2008) wählen den Begriff *frequentistische Model-Averaging Schätzung*, um ihr Verfahren von der *bayesianischen Model-Averaging Schätzung* abzugrenzen. In der bayesianischen Literatur sind Model-Averaging Schätzer bereits seit Langem sehr populär, eine Einführung in dieses Verfahren findet sich in Hoeting et al. (1999). Gemäß Hoeting et al. (1999) berücksichtigen bayesianische Model-Averaging Schätzer die zusätzliche Variabilität der Schätzer, die durch die Modellunsicherheit resultiert.

Um einen bayesianischen Model-Averaging Schätzer zu bestimmen wird die A-posteriori Verteilung des Zielparameters μ für gegebene Daten y_n mittels

$$\mathbb{P}(\mu | y_n) = \sum_{s=1}^r \mathbb{P}(\mu | f_s, y_n) \mathbb{P}(f_s | y_n) \quad (2.25)$$

bestimmt. Dabei ist $\mathbb{P}(\mu | f_s, y_n)$ die A-posteriori Wahrscheinlichkeit für den Parameter μ für gegebene Daten y_n und ein gegebenes Modell f_s . Diese wird mit der A-posteriori Wahrscheinlichkeit $\mathbb{P}(f_s | y_n)$ für das Modell f_s für gegebene Daten y_n gewichtet. Letztere kann berechnet werden über

$$\mathbb{P}(f_s | y_n) = \frac{\mathbb{P}(y_n | f_s) \mathbb{P}(f_s)}{\sum_{s=1}^r \mathbb{P}(y_n | f_s) \mathbb{P}(f_s)}, \quad (2.26)$$

wobei $\mathbb{P}(f_s)$ die A-priori Wahrscheinlichkeit für Modell f_s bezeichnet. Enthält das Modell f_s unbekannte Parameter θ_s , so müssen für diese ebenfalls A-priori Wahr-

scheinlichkeiten $\mathbb{P}(\theta_s | f_s)$ spezifiziert werden. Dann erhalten wir für die Wahrscheinlichkeit $\mathbb{P}(y_n | f_s)$ aus (2.26) mittels Integration

$$\mathbb{P}(y_n | f_s) = \int \mathbb{P}(y_n | \theta_s, f_s) \mathbb{P}(\theta_s | f_s) d\theta_s, \quad (2.27)$$

wobei $\mathbb{P}(y_n | \theta_s, f_s)$ die Likelihood der Daten ist.

Ist die A-posteriori Wahrscheinlichkeit $\mathbb{P}(f_s | y_n)$ bestimmt, so ergibt sich der Model-Averaging Schätzer als gewichtetes Mittel der A-posteriori Erwartungswerte

$\mathbb{E}(\mu | y_n, f_s)$:

$$\hat{\mu} = \mathbb{E}(\mu | y_n) = \sum_{s=1}^r \mathbb{E}(\mu | y_n, f_s) P(f_s | y_n). \quad (2.28)$$

Hoeting et al. (1999) illustrieren anhand zweier Beispiele, dass diese Art der Schätzung präzisere Vorhersagen ermöglicht als die Schätzung in einem einzelnen Modell. Um jedoch einen Schätzer der Form (2.28) zu bestimmen, müssen zunächst A-priori Wahrscheinlichkeiten für die Modelle f_s sowie ihre Parameter θ_s bestimmt werden. Zudem ist die Berechnung der Wahrscheinlichkeit in (2.25) mitunter sehr aufwändig, da eine große Anzahl Terme auftreten kann und das Integral in (2.27) für gewöhnlich mittels Markov-Chain-Monte-Carlo-Verfahren bestimmt werden muss (Hoeting et al., 1999). Da zudem Kapetanios et al. (2008) in Simulationen demonstrieren, dass die frequentistischen Model-Averaging Schätzer ebenso gute Ergebnisse liefern wie ihre bayesianischen Pendanten, konzentriert sich diese Arbeit auf frequentistische Ansätze.

Kapitel 3

Model-Averaging unter lokalen Alternativen

Das Ziel dieser Arbeit ist die Bestimmung optimaler Versuchspläne zur Model-Averaging Schätzung eines Zielparameters μ . Um uns diesem Ziel zu nähern, beschränken wir uns in diesem Kapitel zunächst auf eine spezielle Struktur der Kandidatenmodelle f_1, \dots, f_r , welche in Hjort und Claeskens (2003) verwendet wird, um die asymptotische Verteilung von Model-Averaging Schätzern der Form (2.22) zu beschreiben. Diese Struktur beschreiben wir in Abschnitt 3.1 und erläutern die resultierende asymptotische Verteilung des Model-Averaging Schätzers. Basierend darauf entwickeln wir in Abschnitt 3.2 Kriterien, welche optimale Versuchspläne zur Model-Averaging Schätzung eines Zielparameters μ charakterisieren und leiten für diese notwendige Bedingungen her. Zwei Beispiele illustrieren die Theorie in Abschnitt 3.3 und wir untersuchen die Eigenschaften der resultierenden Versuchspläne anhand von Simulationsstudien. In Abschnitt 3.4 schlagen wir eine adaptive Vorgehensweise vor, in welcher die Modellgewichte anhand der Daten angepasst werden. Beweise zu den vorgestellten theoretischen Resultaten finden sich in Abschnitt 3.5. In Abschnitt 3.6 präsentieren wir weitere Simulationsergebnisse, welche die Beobachtungen aus Abschnitt 3.3 ergänzen. Die in diesem Kapitel vorgestellten Ergebnisse finden sich in Teilen auch in Alhorn et al. (2019a).

3.1 Das Konzept der lokalen Alternativen

Hjort und Claeskens (2003) verwenden das hier vorgestellte Konzept der lokalen Alternativen, um die asymptotische Verteilung von Model-Averaging Schätzern zu bestimmen. Wir erläutern im Folgenden, welche Annahmen sie zusätzlich zu den For-

derungen (A1)-(A6) aus Kapitel 2 stellen. Während in Kapitel 2 das wahre Modell nicht unter den Kandidatenmodellen sein muss, gehen Hjort und Claeskens (2003) davon aus, dass das wahre Modell eine parametrische Dichte g besitzt, welche in der Menge \mathcal{S} von parametrischen Dichten von Kandidatenmodellen enthalten ist. Grundlegend bei diesem Konzept ist die Existenz eines größten Modells mit parametrischer Dichte $f_{\text{wide}}(y|x, \theta, \gamma)$, bei dem die Form der Dichte bekannt ist, die Parameter $\theta = (\theta_1, \dots, \theta_p) \in \Theta$ und $\gamma = (\gamma_1, \dots, \gamma_q) \in \Gamma$ jedoch unbekannt sind. Wir betrachten dabei einen kompakten Parameterraum $\Theta \times \Gamma \subset \mathbb{R}^p \times \mathbb{R}^q$. Diese Notation weicht von der Notation aus Kapitel 2 ab, in welchem die wahre Verteilung lediglich den Parameter $\theta \in \Theta$ enthält. Der zusätzliche Parameter γ dient Hjort und Claeskens (2003) zur Konstruktion der übrigen Dichten, die in \mathcal{S} enthalten sind. So erhalten sie beispielsweise das kleinste in \mathcal{S} enthaltene Modell, indem sie den Parameter γ auf einen zuvor festgelegten (bekannten) Parameterwert $\gamma_0 \in \Gamma$ fixieren. Sie wählen also als Dichte $f_{\text{narrow}}(y|x, \theta) = f_{\text{wide}}(y|x, \theta, \gamma_0)$. Weitere Kandidatenmodelle ergeben sich dann als Submodelle zwischen dem größten und dem kleinsten Modell. Diese erhalten wir, indem wir für eine Teilmenge $S \subset \{1, \dots, q\}$ von Indizes mit Kardinalität $|S|$ die Projektionsmatrizen $\pi_S \in \mathbb{R}^{|S| \times q}$ und $\pi_{S^c} \in \mathbb{R}^{|S^c| \times q}$ einführen, welche einen q -dimensionalen Vektor auf seine Komponenten abbilden, deren Index in S beziehungsweise S^c enthalten ist. Für die entsprechenden Projektionen von γ nutzen wir im Folgenden die Abkürzungen $\gamma_S = \pi_S \gamma$ und $\gamma_{0,S^c} = \pi_{S^c} \gamma_0$. Damit ist dann die Dichte von dem Kandidatenmodell mit den Komponenten aus S gegeben durch

$$f_S(y|x, \theta, \gamma_S) = f_{\text{wide}}(y|x, \theta, \gamma_S, \gamma_{0,S^c}). \quad (3.1)$$

Das bedeutet also, dass in diesem Modell die Komponenten von γ_S unbekannt und somit zu schätzen sind, während die Komponenten von γ mit Index aus $S^c = \{1, \dots, q\} \setminus S$ durch die entsprechenden Komponenten aus γ_0 ersetzt werden. Dieser Ansatz ermöglicht es, gewisse Parameter in jedem Kandidatenmodell zu inkludieren. Er stellt jedoch keine Beschränkung dar, da auch sämtliche Parameter des Modells in γ enthalten sein können.

In dieses Konzept lassen sich auch leicht die oben definierten Dichten einbetten, nämlich ist $f_{\text{narrow}} = f_{\emptyset}$ und $f_{\text{wide}} = f_{\{1, \dots, q\}}$. Auf diese Weise ist es möglich 2^q verschiedene Dichten zu konstruieren. Es mag jedoch Situationen geben, in denen nicht alle Submodelle von Interesse sind, sodass wir lediglich r konkurrierende Kandidatenmodelle mit Indexmenge $S_1, \dots, S_r \subset \{1, \dots, q\}$ betrachten. Um die Notation zu

vereinfachen, schreiben wir im Folgenden analog zu Kapitel 2 für f_{S_i} auch f_i , sodass die Menge \mathcal{S} der Kandidatenmodelle durch

$$\mathcal{S} = \{f_1(y|x, \theta, \gamma_{S_1}), \dots, f_r(y|x, \theta, \gamma_{S_r})\}$$

gegeben ist. Die Bezeichnung *lokale Alternativen* für das hier beschriebene Konzept beruht auf den Annahmen von Hjort und Claeskens (2003) an die wahre Dichte. Diese lässt sich schreiben als

$$g_n(y|x) = f_{\text{wide}}(y|x, \theta_0, \gamma_0 + \frac{\delta}{\sqrt{n}}), \quad (3.2)$$

sodass die wahren Parameterwerte also $\theta_0 \in \Theta$ und $\gamma_0 + \frac{\delta}{\sqrt{n}} \in \Gamma$ sind. Es handelt sich hierbei um die Dichte des größten Modells mit einem mit der Stichprobengröße n variierenden Wert von γ . Mit wachsendem n nähert sich die wahre Dichte der Dichte des kleinsten Modells an.

Beispiel 3.1. Wir betrachten in diesem Beispiel ein Regressionsmodell analog zu (2.1), wobei wir hier die Notation an das oben erläuterte Konzept anpassen. Dazu gehen wir davon aus, dass wir normalverteilte Fehler haben, sodass $f_{\text{wide}}(y|x, \theta, \gamma) = f_4(y|x, \theta, \gamma)$ die Dichte einer Normalverteilung ist mit Varianz σ^2 und Erwartungswertfunktion

$$\eta_4(x, \vartheta, \gamma) = \gamma_1 + \vartheta_1 \frac{x^{\gamma_2}}{x^{\gamma_2} + \vartheta_2^{\gamma_2}}, \quad (3.3)$$

mit $\theta^T = (\sigma^2, \vartheta_1, \vartheta_2)$ und $\gamma^T = (\gamma_1, \gamma_2)$. Bei diesem Modell handelt es sich um das weit verbreitete *sigmoidale Emax-Modell*. Es findet Anwendung bei der Modellierung von biochemischen oder pharmakologischen Konzentrations-Wirkungs-Beziehungen (siehe beispielsweise Goutelle et al. (2008) für einen Überblick). Vor allem in der Medikamentenentwicklung ist dieses Modell beliebt, um Dosis-Wirkungs-Beziehungen zu beschreiben (siehe beispielsweise MacDougall (2006)). Eine grafische Darstellung der Erwartungswertfunktion (3.3) findet sich in Abbildung 3.1. In dieser verwenden wir feste Parameter $\vartheta_1 = 1.81$, $\vartheta_2 = 0.79$ und $\gamma_1 = 0$, während für γ_2 jeweils unterschiedliche Werte abgetragen sind. In der Terminologie, die im Zusammenhang von Dosis-Wirkungs-Beziehungen gebräuchlich ist, lassen sich die Parameter wie folgt interpretieren: γ_1 stellt den Placebo-Effekt eines Medikamentes dar, während ϑ_1 der asymptotische maximale Effekt ist (relativ zum Placebo-Effekt). Die Dosis, bei der die Hälfte des maximalen Effektes erreicht wird, kann an ϑ_2 abgelesen werden. Es zeigt sich beispielsweise in Abbildung 3.1 für alle Kurven, dass im Punkt $\vartheta_2 = 0.79$ die Hälfte des maximalen Effektes $\vartheta_1 = 1.81$ erzielt wird. Der sogenannte

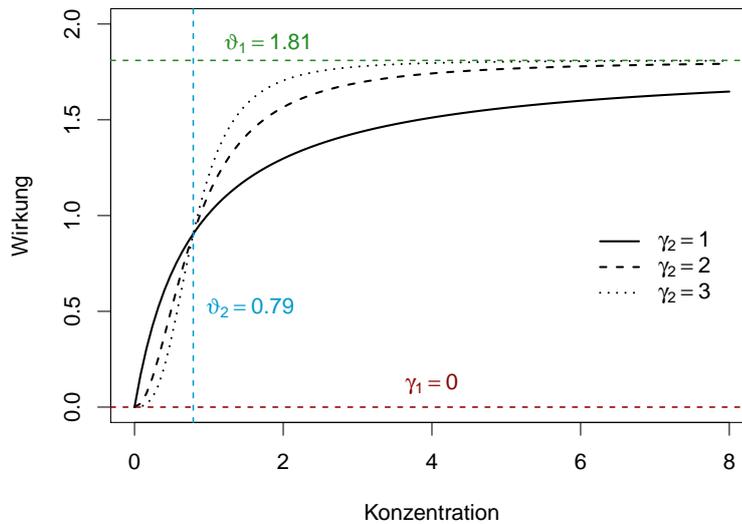


Abbildung 3.1: Erwartungswertfunktion η_4 des sigmoidalen Emax-Modells aus (3.3) für $\vartheta_1 = 1.81$, $\vartheta_2 = 0.79$, $\gamma_1 = 0$ und $\gamma_2 \in \{1, 2, 3\}$.

Hill-Parameter γ_2 bestimmt die Steilheit der Kurve. Je größer dieser Parameterwert ist, desto ausgeprägter ist die S-Form der Kurve (siehe Abbildung 3.1).

Ausgehend von dem größten Modell können weitere Kandidatenmodelle konstruiert werden. Dabei sei der Parameter θ in jedem der Kandidatenmodelle enthalten, während die Komponenten von γ fixiert werden können auf $\gamma_0 = (0, 1)^T$. Dadurch besitzt das kleinste Modell $f_{\text{narrow}}(y|x, \theta) = f_1(y|x, \theta)$ eine Normalverteilung mit Varianz σ^2 und Erwartungswertfunktion

$$\eta_1(x, \vartheta) = \frac{\vartheta_1 x}{x + \vartheta_2}. \quad (3.4)$$

Hierdurch entspricht η_1 der *Michaelis-Menten-Gleichung*. Diese ist in der Biochemie gebräuchlich, um Enzymkinetiken zu beschreiben, bei denen Enzyme an Substrate binden und zu Produkten umsetzen (siehe beispielsweise Cornish-Bowden (2012)). Die weiteren Kandidatenmodelle ergeben sich durch Fixierung von entweder $\gamma_1 = 0$ oder $\gamma_2 = 1$, sodass die resultierenden Kandidatenmodelle die Dichten von Normalverteilungen mit Varianz σ^2 und Erwartungswertfunktion

$$\eta_2(x, \vartheta, (0, \gamma_2)) = \vartheta_1 \frac{x^{\gamma_2}}{x^{\gamma_2} + \vartheta_2^{\gamma_2}} ; \quad \eta_3(x, \vartheta, (\gamma_1, 1)) = \gamma_1 + \vartheta_1 \frac{x}{x + \vartheta_2} \quad (3.5)$$

sind. Dabei erhalten wir mit η_3 das *Emax-Modell*, welches zur Unterscheidung vom sigmoidalen Emax-Modell auch hyperbolisches Emax-Modell genannt wird und ebenfalls Anwendung in der Modellierung von Dosis-Wirkungs-Beziehungen findet (siehe

beispielsweise Holford und Sheiner (1981) oder MacDougall (2006)). Abschließend ergibt sich unter der Annahme, dass wir lokale Alternativen vorliegen haben, die wahre Dichte $g_n(y|x)$ als Dichte einer Normalverteilung mit Mittelwert

$$\eta_{\text{true},n}(x) = \frac{\delta_1}{\sqrt{n}} + \vartheta_1 \frac{x^{1+\delta_2/\sqrt{n}}}{x^{1+\delta_2/\sqrt{n}} + \vartheta_2^{1+\delta_2/\sqrt{n}}}$$

und Varianz σ^2 .

Wie in Kapitel 2 beschrieben soll mit den vorliegenden Kandidatenmodellen ein Fokusparameter μ geschätzt werden. In obiger Notation hängt dieser Fokusparameter von den Parametern θ und γ ab, sodass wir auch schreiben $\mu(\theta, \gamma)$. Wir setzen voraus, dass $\mu : \Theta \times \Gamma \rightarrow \mathbb{R}$ eine bezüglich der Parameter (θ, γ) differenzierbare Funktion ist. Wir können dann den Schätzer des Fokusparameters aus (2.12) unter Verwendung des Modells f_s in der hier verwendeten Notation schreiben als

$$\hat{\mu}_s = \mu(\hat{\theta}, \hat{\gamma}_{S_s}, \gamma_{0,S_s^c}), s = 1, \dots, r, \quad (3.6)$$

wobei $(\hat{\theta}, \hat{\gamma}_{S_s}) \in \mathbb{R}^{p+|S_s|}$ der Maximum-Likelihood-Schätzer der Parameter im Modell f_s ist.

Beispiel 3.2. (Fortsetzung von Beispiel 3.1)

In Dosis-Wirkungs-Studien sind typische Fokusparameter μ die Fläche unter der Kurve (AUC)

$$\mu(\theta, \gamma) = \int_{\mathcal{C}} \eta(x, \vartheta, \gamma) dx \quad (3.7)$$

für einen gegebenen Bereich $\mathcal{C} \subset \mathbb{R}$ oder für gegebenes $\alpha \in (0, 1)$ und eine monotone Funktion η das „Quantil“ definiert über

$$\mu(\theta, \gamma) = \inf \left\{ x \in \mathcal{X} \mid \frac{\eta(x, \vartheta, \gamma) - \eta(x_l, \vartheta, \gamma)}{\eta(x_u, \vartheta, \gamma) - \eta(x_l, \vartheta, \gamma)} \geq \alpha \right\}. \quad (3.8)$$

Der Wert, der in (3.8) definiert ist, ist die ED_α . Dies ist die effektive Dosis, bei der $100 \times \alpha\%$ des maximalen Effekts im Versuchsraum $\mathcal{X} = [x_l, x_u]$ erreicht wird (siehe MacDougall (2006) oder Bretz et al. (2008)).

Auch wenn die Verwendung der oben beschriebenen Annahmen in Teilen sehr restriktiv ist, so hat sie den Vorteil, dass sie zu handlichen Ausdrücken für die asymptotische Verteilung der Schätzer in (3.6) führt. Zur Herleitung dieser gehen wir davon aus, dass die Annahmen (A1)-(A6) erfüllt sind, die typisch im Kontext der

Maximum-Likelihood-Theorie sind (siehe Kapitel 2). Wir benötigen zudem eine weitere Annahme aus der klassischen Maximum-Likelihood Theorie, nämlich dass das Integral

$$\int f_{\text{wide}}(y | x, \theta, \gamma) d\nu(y)$$

unter dem Integralzeichen zweimal differenzierbar ist (Lehmann, 1983, Kapitel 6). In Anlehnung an die Notation aus Hjort und Claeskens (2003) schreiben wir im Folgenden für die Matrix aus (2.18) auch $J_s(\xi, \theta_0, \gamma_{0,S_s})$. Hierbei handelt es sich um die *Informationsmatrix* im Kandidatenmodell f_s , genauer

$$J_s(\xi, \theta_0, \gamma_{0,S_s}) = \int_{\mathcal{X}} \int \frac{\left(\frac{\partial}{\partial(\theta, \gamma_{S_s})} f_s(y|x, \theta_0, \gamma_{0,S_s})\right) \left(\frac{\partial}{\partial(\theta, \gamma_{S_s})} f_s(y|x, \theta_0, \gamma_{0,S_s})\right)^T}{f_s(y|x, \theta_0, \gamma_{0,S_s})} dy \xi(dx). \quad (3.9)$$

Satz 3.3. *Unter der Annahme der lokalen Alternativen sowie den genannten Regularitätsbedingungen gilt für den Schätzer des Zielparameters $\hat{\mu}_s$ im Modell f_s*

$$\sqrt{n} \left(\hat{\mu}_s - \mu(\theta_0, \gamma_0 + \frac{\delta}{\sqrt{n}}) \right) \xrightarrow{\mathcal{D}} \Lambda_s \sim \mathcal{N}(\nu_s(\xi), \tau_s^2(\xi)). \quad (3.10)$$

Dabei ist $\mathcal{N}(\nu_s(\xi), \tau_s^2(\xi))$ eine Normalverteilung mit Varianz

$$\tau_s^2(\xi) = \tau_s^2(\xi, \theta_0, \gamma_0) = c_s^T J_s^{-1}(\xi, \theta_0, \gamma_{0,S_s}) c_s,$$

wobei c_s der Gradient der Zielfunktion μ bezüglich (θ, γ_{S_s}) ist, das bedeutet

$$c_s = c_s(\theta_0, \gamma_{0,S_s}) = \frac{\partial}{\partial(\theta, \gamma_{S_s})} \mu(\theta, \gamma_{S_s}, \gamma_{0,S_s^c}) \Big|_{(\theta, \gamma_{S_s}) = (\theta_0, \gamma_{0,S_s})}. \quad (3.11)$$

Weiterhin ist J_s die Informationsmatrix im Kandidatenmodell f_s aus (3.9). Der Mittelwert in (3.10) lässt sich darstellen als

$$\nu_s(\xi) = \nu_s(\xi, \delta, \theta_0, \gamma_0) = c^T L_s(\xi, \theta_0, \gamma_0) \delta, \quad (3.12)$$

wobei

$$c = c(\theta_0, \gamma_0) = \frac{\partial}{\partial(\theta, \gamma)} \mu(\theta, \gamma) \Big|_{(\theta, \gamma) = (\theta_0, \gamma_0)} \quad (3.13)$$

der Gradient von μ im größten Modell ist und wir die Matrix L_s definieren als

$$L_s(\xi, \theta_0, \gamma_0) = \left(P_s^T J_s^{-1}(\xi, \theta_0, \gamma_{0,S_s}) P_s J(\xi, \theta_0, \gamma_0) - I_{(p+q) \times (p+q)} \right) \begin{pmatrix} 0_{p \times q} \\ I_{q \times q} \end{pmatrix}. \quad (3.14)$$

Dabei ist $J(\xi, \theta_0, \gamma_0)$ die Informationsmatrix im größten Modell f_{wide} und P_s ist gegeben durch

$$P_s = \begin{pmatrix} I_{p \times p} & 0 \\ 0 & \pi_{S_s} \end{pmatrix}. \quad (3.15)$$

Bemerkung. Im größten Modell ergibt sich $\nu_s(\xi) = 0$, sodass hier die Schätzung unverzerrt ist. Jedoch kann die Schätzung vieler Parameter zu einer großen Varianz führen. Somit können wir die Varianz der Parameterschätzung verringern, wenn wir nicht sämtliche Parameter schätzen. Dies führt allerdings zu einer Verzerrung, sodass wir bei der Modellwahl stets zwischen Verzerrung und Varianz abwägen müssen. Claeskens und Hjort (2008, Kapitel 5) erläutern diesen Konflikt und geben für gewisse Situationen Kriterien an, ob die Schätzung im großen oder im kleinen Modell genauer ist (gemessen am mittleren quadratischen Fehler). Basierend auf der Erkenntnis, dass die Wahl des Modells insbesondere vom Fokusparameter μ abhängt, entwickeln Claeskens und Hjort (2003) das *Fokussierte Informationskriterium (FIC)*. Im Gegensatz zum AIC oder BIC wählt dieses Kriterium das Modell, das sich zur Schätzung des Zielparameters am besten eignet. Dabei ist das am besten geeignete Modell dasjenige, in dem der Schätzer $\hat{\mu}_s$ aus (2.12) beziehungsweise (3.6) den geringsten asymptotischen mittleren quadratischen Fehler besitzt, welchen sie mit den Resultaten aus Satz 3.3 bestimmen.

Soll ein Model-Averaging Schätzer (2.22) verwendet werden, können wir basierend auf Satz 3.3 die asymptotische Verteilung dessen angeben.

Satz 3.4. *Unter der Annahme lokaler Alternativen (3.2) und den genannten Regularitätsbedingungen gilt für den Model-Averaging Schätzer $\hat{\mu}_{max}$ aus (2.22) bei festen Gewichten w_1, \dots, w_r :*

$$\sqrt{n} \left(\hat{\mu}_{max} - \mu(\theta_0, \gamma_0 + \frac{\delta}{\sqrt{n}}) \right) \xrightarrow{\mathcal{D}} \sum_{s=1}^r w_s \Lambda_s \sim \mathcal{N}(\nu(\xi, \delta, \theta_0, \gamma_0), \tau^2(\xi, \theta_0, \gamma_0)). \quad (3.16)$$

Dabei sind Erwartungswert und Varianz der resultierenden Normalverteilung gegeben durch

$$\nu(\xi, \delta, \theta_0, \gamma_0) = \sum_{s=1}^r w_s \nu_s(\xi, \delta, \theta_0, \gamma_0), \quad (3.17)$$

$$\tau^2(\xi, \theta_0, \gamma_0) = \sum_{s=1}^r \sum_{t=1}^r w_s w_t h_s^T(\xi) J(\xi, \theta_0, \gamma_0) h_t(\xi). \quad (3.18)$$

Der Ausdruck $\nu_s(\xi, \delta, \theta_0, \gamma_0)$ ist in (3.12) gegeben und $J(\xi, \theta_0, \gamma_0)$ ist wie in Satz 3.3 die Informationsmatrix im größten Modell f_{wide} . Weiterhin definieren wir den Vektor $h_s(\xi)$ als

$$h_s(\xi) = P_s^T J_s^{-1}(\xi, \theta_0, \gamma_{0,s}) c_s, \quad (3.19)$$

wobei c_s , J_s und P_s gemäß (3.11), (3.9) und (3.15) verwendet werden.

Bemerkung. Hjort und Claeskens (2003) verzichten auf die Annahme fester Gewichte w_1, \dots, w_r und nutzen stattdessen zufällige Gewichte wie beispielsweise in (2.23) und (2.24). Hierdurch lassen sich die Verteilung von Model-Averaging Schätzern und von Schätzern nach Modellselektion in einem einheitlichen Rahmen darstellen. Jedoch ist die resultierende asymptotische Verteilung bei der Verwendung von zufälligen Gewichten eine nichtlineare Kombination von zwei Zufallsvariablen, die nicht notwendigerweise einer Normalverteilung folgt. Ist der Parameter γ eindimensional, so kann die asymptotische Verteilung über numerische Integration bestimmt werden. Für höherdimensionale Parametervektoren ist dies nicht ohne Weiteres möglich, sodass Claeskens und Hjort (2008, S. 197) eine rechenintensive Näherung der Verteilung mittels Simulationen vorschlagen. Diese Arbeit verfolgt diesen Ansatz aufgrund der aufwändigen Bestimmung der Verteilung nicht weiter. Zudem beobachten wir in den durchgeführten Simulationen in Abschnitt 3.3, dass in vielen Szenarien die Schätzung basierend auf einer festen Gewichtung w_1, \dots, w_r geringere mittlere quadratische Fehler aufweist als die Schätzung basierend auf zufälligen Gewichten.

3.2 Versuchsplanungskriterien

Wir haben in Abschnitt 3.1 die asymptotische Verteilung von Model-Averaging Schätzern $\hat{\mu}_{\text{mav}}$ aus (2.22) im Kontext lokaler Alternativen hergeleitet. Satz 3.4 zeigt, dass diese von den betrachteten Modellen sowie deren unbekanntem Parametern aber auch maßgeblich vom verwendeten Versuchsplan ξ abhängt. Model-Averaging Schätzer sind gemäß (3.16) im Allgemeinen nicht erwartungstreu für den Fokusparameter $\mu(\theta_0, \gamma_0 + \delta/\sqrt{n})$, sodass ein guter Versuchsplan sowohl die Verzerrung als auch die Varianz dieses Schätzers minimieren sollte. Somit betrachten wir im Folgenden den mittleren quadratischen Fehler (MSE) der asymptotischen Verteilung des Schätzers, welcher sich mit Satz 3.4 ergibt als

$$\Phi_{\text{mav}}(\xi, w, \delta, \theta_0, \gamma_0) = \nu^2(\xi, \delta, \theta_0, \gamma_0) + \tau^2(\xi, \theta_0, \gamma_0) \approx n \cdot \text{MSE}(\hat{\mu}_{\text{mav}}), \quad (3.20)$$

wobei die Varianz $\tau^2(\xi, \theta_0, \gamma_0)$ und die Verzerrung $\nu(\xi, \delta, \theta_0, \gamma_0)$ in (3.17) und (3.18) angegeben sind.

Definition 3.5. Für eine endliche Menge \mathcal{S} von Kandidatenmodellen f_s , welche gemäß (3.1) konstruiert ist und feste Gewichte w_s ($s = 1, \dots, r$) heißt ein Versuchsplan ξ^* *lokal optimaler Versuchsplan zur Model-Averaging Schätzung des Parameters μ* ,

falls er die Funktion $\Phi_{\text{mav}}(\xi, w, \delta, \theta_0, \gamma_0)$ in (3.20) in der Klasse aller approximativen Versuchspläne auf \mathcal{X} minimiert, das heißt falls

$$\Phi_{\text{mav}}(\xi^*, w, \delta, \theta_0, \gamma_0) = \min_{\xi \in \Xi} \Phi_{\text{mav}}(\xi, w, \delta, \theta_0, \gamma_0).$$

Wir verwenden die Bezeichnung „lokal“ im Sinne von Chernoff (1953), da lokal optimale Versuchspläne zur Model-Averaging Schätzung des Parameters μ zwar die Unsicherheit bezüglich des Modells f_s berücksichtigen, jedoch Vorwissen über die Parameter θ_0, γ_0 und δ nötig ist, um sie zu bestimmen. Dieses mag in gewissen Situationen zwar vorliegen, etwa durch vorherige Studien (siehe beispielsweise Dette et al. (2008) oder Bretz et al. (2010)), oftmals ist es jedoch notwendig Versuchsplanungskriterien zu verwenden, die weniger Vorwissen über die Modellparameter benötigen. Dazu zählen sequenzielle, bayesianische oder standardisierte Maximin-Kriterien. Bei sequenziellen Versuchsplänen wird ein Parameterschätzer zur Bestimmung eines optimalen Versuchsplans verwendet, welcher mit Ergebnissen aus neuen Experimenten jeweils angepasst wird, sodass basierend auf dem verbesserten Schätzer erneut Versuche geplant werden können. Die Bestimmung der Verteilung solcher Verfahren ist jedoch wesentlich komplexer, da die Zufallsvariablen Y_{ij} nicht mehr unabhängig sind (siehe beispielsweise Silvey (1980, Kapitel 7)). Standardisierte Maximin-Kriterien basieren auf einem *Worst-Case-Szenario*, bei dem ein optimaler Versuchsplan die minimale Effizienz maximiert (siehe Silvey (1980, S. 59)). Wir konzentrieren uns in dieser Arbeit auf Bayes-optimale Versuchspläne (siehe Chaloner und Verdinelli (1995) für eine Übersichtsarbeit), bei welchen wir die Unsicherheit über die unbekannt Parameter mit einer A-priori Verteilung π auf $\Theta \times \Gamma$ quantifizieren. Basierend auf dieser definieren wir ein neues Versuchsplanungskriterium.

Definition 3.6. Unter den Annahmen aus Definition 3.5 heißt ein Versuchsplan ξ^* *Bayes-optimal zur Model-Averaging Schätzung des Parameters μ hinsichtlich der A-priori Verteilung π* , falls er die Funktion

$$\Phi_{\text{mav}}^{\pi}(\xi) = \int_{\Theta \times \Gamma} \Phi_{\text{mav}}(\xi, w, \delta, \theta, \gamma) \pi(d\theta, d\gamma) \quad (3.21)$$

minimiert, das heißt, falls

$$\Phi_{\text{mav}}^{\pi}(\xi^*) = \min_{\xi \in \Xi} \Phi_{\text{mav}}^{\pi}(\xi).$$

Wir gehen dabei stets davon aus, dass das Integral in (3.21) existiert.

Ein Bayes-optimaler Versuchsplan minimiert also den unter der A-priori Verteilung π erwarteten asymptotischen mittleren quadratischen Fehler des Model-Averaging

Schätzers. Wir beschränken uns in dieser Arbeit auf eine A-priori Verteilung für die Parameter θ und γ . In Anwendungen ist die hier vorgestellte Theorie jedoch leicht um eine A-priori Verteilung für den Parameter δ erweiterbar, sodass Unsicherheiten bezüglich dieses Parameters ebenfalls berücksichtigt werden können.

Ein Bayes-optimaler Versuchsplan zur Model-Averaging Schätzung des Parameters μ hängt zwar nicht mehr von den Parametern θ und γ ab, dafür jedoch von der Wahl der A-priori Verteilung π . In einigen Anwendungen kann zur Bestimmung dieser Vorwissen verwendet werden. Dies ist typischerweise in Dosisfindungsstudien der Fall, in denen Informationen aus der pharmazeutischen Entwicklung oder einer bereits durchgeführten Phase I Studie vorliegen (siehe beispielsweise Dette et al. (2008)). Zudem ist die Verwendung nichtinformativer A-priori Verteilungen möglich. Die durch den Bayes-optimalen Versuchsplan erzielte Schätzgenauigkeit hängt von der Wahl der A-priori Verteilung ab. Insbesondere ist es von zentraler Bedeutung, wie gut die A-priori Verteilung die Wahrheit widerspiegelt. In Simulationen zeigt sich, dass die Schätzungen basierend auf lokal optimalen Versuchsplänen zur Model-Averaging Schätzung des Parameters μ geringere mittlere quadratische Fehler aufweisen als die Schätzungen basierend auf Bayes-optimalen Versuchsplänen zur Model-Averaging Schätzung des Parameters μ , sofern die zuvor getätigten Parameterannahmen sehr präzise sind. Sind die Parameter jedoch falsch spezifiziert, so weisen Bayes-optimale Versuchspläne im Allgemeinen bessere Ergebnisse (hinsichtlich der Schätzgenauigkeit) auf.

Die Berechnung der Kriterien $\Phi_{\text{mav}}(\xi)$ und $\Phi_{\text{mav}}^\pi(\xi)$ liefert sehr komplexe Ausdrücke, deren analytische Minimierung im Allgemeinen nicht möglich ist. Daher bestimmen wir optimale Versuchspläne numerisch, Beispiele hierzu finden sich in Abschnitt 3.3. Um zu überprüfen, ob ein numerisch bestimmter Versuchsplan Bayes-optimal zur Model-Averaging Schätzung des Parameters μ sein kann, geben wir nun eine notwendige Bedingung analog zu Satz 2.5 an.

Satz 3.7. *Ist ein Versuchsplan ξ^* Bayes-optimal zur Model-Averaging Schätzung des Parameters μ hinsichtlich der A-priori-Verteilung π , dann gilt für alle $x \in \mathcal{X}$*

$$d_\pi(x, \xi^*) = \int_{\Theta \times \Gamma} -2\nu(\xi^*, \delta, \theta, \gamma) D_1(\xi^*, x, \delta, \theta, \gamma) - D_2(\xi^*, x, \theta, \gamma) \pi(d\theta, d\gamma) \leq 0. \quad (3.22)$$

Dabei ist $\nu(\xi^*, \delta, \theta, \gamma)$ in (3.17) definiert und die Ableitungen D_1 und D_2 sind gegeben durch

$$D_1(\xi^*, x, \delta, \theta_0, \gamma_0) = \sum_{s=1}^r w_s c^T P_s^T J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) \left(P_s J(\xi_x, \theta_0, \gamma_0) - J_s(\xi_x, \theta_0, \gamma_{0,S_s}) J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) P_s J(\xi^*, \theta_0, \gamma_0) \right) \begin{pmatrix} 0_p \\ \delta \end{pmatrix}, \quad (3.23)$$

$$D_2(\xi^*, x, \theta_0, \gamma_0) = \sum_{s,t=1}^r w_s w_t (h_s^T(\xi^*) \{ J(\xi^*, \theta_0, \gamma_0) + J(\xi_x, \theta_0, \gamma_0) \} h_t(\xi^*) - 2\tilde{h}_s^T(\xi^*, \xi_x) J(\xi^*, \theta_0, \gamma_0) h_t(\xi^*)), \quad (3.24)$$

wobei die Vektoren c und $h_s(\xi)$ in (3.13) und (3.19) gegeben sind und die Matrizen $J(\xi^*, \theta_0, \gamma_0)$, $J_s(\xi_x, \theta_0, \gamma_{0,S_s})$ und P_s in Satz 3.3, (3.9) und (3.15) definiert sind. Der Vektor $\tilde{h}_s(\xi^*, \xi)$ sei definiert als

$$\tilde{h}_s(\xi^*, \xi) = P_s^T J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) J_s(\xi, \theta_0, \gamma_{0,S_s}) J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) c_s, \quad (3.25)$$

wobei c_s gemäß (3.11) zu bestimmen ist. Der Versuchsplan ξ_x ist das Einpunktdesign aus (2.6).

Darüber hinaus gilt die Gleichheit in (3.22) für alle Trägerpunkte x von ξ^* .

Wir geben die notwendige Bedingung für das Optimalitätskriterium aus Definition 3.5 nicht explizit an, da dieses ein Spezialfall des Kriteriums in Definition 3.6 ist, wenn wir als A-priori-Verteilung eine Einpunktverteilung in (θ_0, γ_0) wählen. Satz 3.7 kann nun verwendet werden, um zu überprüfen, ob ein numerisch bestimmter Versuchsplan Bayes-optimal zur Model-Averaging Schätzung des Parameters μ sein kann: Erfüllt das numerisch bestimmte Design ξ^* die Ungleichheit in (3.22) *nicht*, so kann er nicht optimal zur Model-Averaging Schätzung des Parameters μ sein und es muss weiter nach einem optimalen Versuchsplan gesucht werden. Es ist zu beachten, dass die Funktionen Φ_{mav} und Φ_{mav}^π nicht konvex sind und somit keine hinreichende Bedingung im Sinne von Satz 2.7 vorliegt.

3.3 Anwendung und Simulation

In diesem Kapitel untersuchen wir das Verhalten von Bayes-optimalen Versuchsplänen zur Model-Averaging Schätzung des Parameters μ . Wir verwenden dazu zwei Szenarien aus Simulationsstudien im Rahmen von Dosis-Wirkungs-Studien (Bornkamp et al., 2007; Schorning et al., 2016) und vergleichen die resultierenden Bayes-optimalen Versuchspläne mit zwei gleichverteilten Versuchsplänen.

In den verwendeten Beispielen berechnen wir die Ableitungen in den Informationsmatrizen J und J_s aus (3.9) sowie dem Vektor c aus (3.13) analytisch, um die numerischen Fehler in der Berechnung des Versuchsplanungskriteriums Φ_{mav}^π aus (3.21) und seiner Richtungsableitung d_π aus (3.22) möglichst klein zu halten. Dennoch kann die numerische Berechnung der Inversen J_s^{-1} zu erheblichen numerischen Fehlern führen, falls die Informationsmatrix schlecht konditioniert ist. Aus diesem Grund ist es zu empfehlen; vor der Bestimmung optimaler Versuchspläne die Konditionszahlen der Matrizen J und J_s sorgfältig zu untersuchen.

Zur Bestimmung von optimalen Versuchsplänen existieren viele Algorithmen (siehe beispielsweise Fedorov und Leonov (2013, Kapitel 3)), von denen viele auf der Richtungsableitung des entsprechenden Versuchsplanungskriteriums basieren. Dadurch können die Algorithmen rechnerisch weniger aufwändig sein als die direkte Optimierung des Kriteriums. Im vorliegenden Fall ist jedoch die Berechnung der Richtungsableitung d_π aus (3.22) nicht nur komplizierter als die Kalkulation des Kriteriums Φ_{mav}^π aus (3.21), sondern auch numerisch instabiler. Aus diesem Grund verzichten wir in dieser Arbeit auf derartige Algorithmen und minimieren das Versuchsplanungskriterium Φ_{mav}^π direkt bezüglich der Trägerpunkte x_1, \dots, x_k und der Gewichte ξ_1, \dots, ξ_k ($i = 1, \dots, k$) mittels des *COBYLA-Algorithmus* (Powell, 1994).

3.3.1 Schätzung der ED_α im sigmoidalen Emax-Modell

Wir betrachten die Situation aus Beispiel 3.1, bei dem wir angenommen haben, dass die zugrunde liegende Dichte die Dichte einer Normalverteilung ist mit Varianz σ^2 und verschiedene Regressionsfunktionen für den Erwartungswert möglich sind. Genauer gesagt enthält die Menge \mathcal{S} der möglichen Kandidatenmodelle $r = 4$ Elemente, die sich durch die verschiedenen Erwartungswertfunktionen in (3.3), (3.4) und (3.5) unterscheiden. Der Zielparameter μ sei nun die $ED_{0.6}$ aus (3.8). Der Versuchsraum ist gewählt als $\mathcal{X} = [0, 8]$ und wir gehen davon aus, dass wir $n = 150$ Beobachtungen mit dem Versuchsplan durchführen können.

Basierend auf diesen Voraussetzungen bestimmen wir einen Bayes-optimalen Versuchsplan zur Model-Averaging Schätzung der $ED_{0.6}$. Zunächst müssen wir also eine A-priori Verteilung für die Parameter spezifizieren. Das sigmoidale Emax-Modell ist linear in den Parametern ϑ_1 und γ_1 , sodass das Optimalitätskriterium nicht von ihnen abhängt und wir kein Vorwissen über diese beiden Parameter benötigen. Für die Parameter (ϑ_2, γ_2) dagegen wählen wir unabhängige A-Priori Verteilungen π_{ϑ_2} und π_{γ_2} , welche gleichverteilt sind auf den Mengen $\{0.79, 1.79, 2.79\}$ und $\{1, 2, 3\}$. Für die Varianz σ^2 wählen wir eine Einpunktverteilung in $\sigma_0^2 = 4.5$. Abschließend wählen

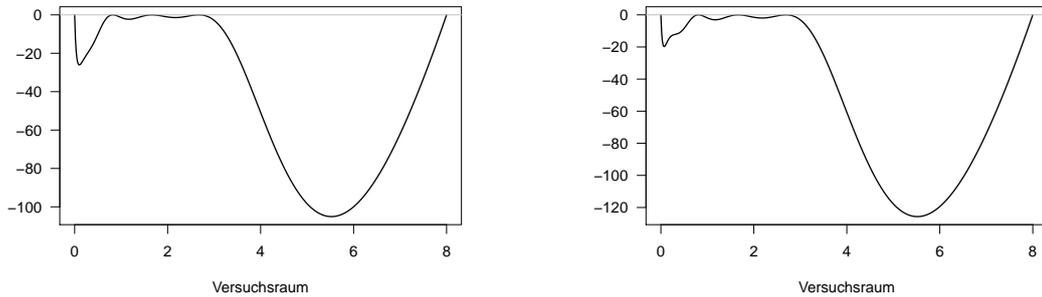


Abbildung 3.2: Die Funktion d_π aus (3.22) ausgewertet für den Versuchsplan ξ_A^* aus (3.26) (linke Grafik) und den Versuchsplan ξ_B^* aus (3.30) (rechte Grafik).

wir in der Annahme der lokalen Alternativen δ so, dass $\delta/\sqrt{n} = \delta/\sqrt{150} = (0.1, 1)^T$ ist.

Weitere festzulegende Parameter sind die Gewichte für den Model-Averaging Schätzer. Hier betrachten wir zunächst eine gleiche Gewichtung aller Modelle, das heißt

$$w_s = 0.25, s = 1, \dots, 4.$$

Dann ist der Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der $ED_{0.6}$ gegeben durch

$$\xi_A^* = \begin{Bmatrix} 0 & 0.819 & 1.665 & 2.669 & 8 \\ 0.105 & 0.138 & 0.199 & 0.273 & 0.285 \end{Bmatrix}. \quad (3.26)$$

Dieser Versuchsplan erfüllt die notwendige Bedingung aus Satz 3.7 (siehe Abbildung 3.2). Wäre diese Bedingung nicht erfüllt, so könnte ξ_A^* nicht der optimale Versuchsplan sein.

Um nun die Eigenschaften der verschiedenen Versuchspläne zur Model-Averaging Schätzung zu untersuchen, führen wir eine Simulationsstudie durch, in der wir zunächst den Bayes-optimalen Versuchsplan (3.26) mit zwei gleichverteilten Versuchsplänen

$$\xi_1 = \begin{Bmatrix} 0 & 2 & 4 & 6 & 8 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{Bmatrix}, \quad (3.27)$$

$$\xi_2 = \begin{Bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \end{Bmatrix} \quad (3.28)$$

vergleichen, die auch in Schorning et al. (2016) und Bornkamp et al. (2007) Verwendung finden. Dabei besitzt ξ_1 die gleiche Anzahl an Trägerpunkten wie (3.26),

während ξ_2 deutlich mehr Trägerpunkte hat. Zusätzlich zu dem Vergleich mit anderen Versuchsplänen, möchten wir untersuchen, wie der Model-Averaging Schätzer mit festen Gewichten abschneidet, wenn wir ihn mit zwei Schätzern, die häufig in der Praxis genutzt werden, vergleichen. Diese Schätzer sind zum einen ein Model-Averaging Schätzer mit zufälligen Gewichten, wobei wir hier die glatte AIC-Gewichtung aus (2.24) nutzen, zum Anderen führen wir eine Modellselektion mittels des AIC durch und schätzen die $ED_{0.6}$ in dem gewählten Modell (siehe etwa Verrier et al. (2014) für eine Anwendung dieser Verfahren in Dosis-Wirkungs Studien). Es ist zu beachten, dass sich auch der Schätzer nach Modellselektion als Model-Averaging Schätzer (2.22) unter Verwendung der Gewichtung in (2.23) auffassen lässt. Zudem untersuchen wir das Verhalten des unverzerrten Schätzers im größten Modell f_{wide} , das sich durch Verwendung von η_4 aus (3.3) ergibt. Auch für diese Schätzer nutzen wir Beobachtungen, die wir gemäß der Versuchspläne ξ_A^* , ξ_1 oder ξ_2 generieren. Um die Versuchspläne nutzen zu können, nutzen wir die in Abschnitt 2.1 beschriebene Rundungsprozedur aus Pukelsheim und Rieder (1992). Für $n = 150$ Beobachtungen bedeutet dies beispielsweise für den Versuchsplan ξ_A^* aus (3.26), dass jeweils $n_1 = 16$, $n_2 = 21$, $n_3 = 30$, $n_4 = 40$ und $n_5 = 43$ Beobachtungen an den Trägerpunkten 0, 0.819, 1.165, 2.669 und 8 zu nehmen sind.

Die hier gezeigten Ergebnisse basieren jeweils auf 1000 Simulationsdurchläufen (im Folgenden bezeichne (l) jeweils das Ergebnis des l -ten Simulationsdurchlaufs) in denen wir jeweils 150 Beobachtungen der Art

$$y_{ij}^{(l)} = \gamma_1 + \vartheta_1 \frac{x_i^{\gamma_2}}{x_i^{\gamma_2} + \vartheta_2^{\gamma_2}} + \sigma \varepsilon_{ij}^{(l)}, i = 1, \dots, k, j = 1, \dots, n_i, \quad (3.29)$$

für die vier verschiedenen Versuchspläne erzeugen. Dabei sind die $\varepsilon_{ij}^{(l)}$ unabhängige, standardnormalverteilte Zufallsvariablen. Wir nutzen in (3.29) verschiedene „wahre“ Parameter $(\vartheta^T, \gamma^T) = (\vartheta_1, \vartheta_2, \gamma_1, \gamma_2)$, die Varianz dagegen wählen wir wie bei der Bestimmung des optimalen Versuchsplans fest als $\sigma^2 = 4.5$.

Wir erläutern nun die Ergebnisse für die Parameterkombination $(\vartheta_1, \vartheta_2) = (1.81, 0.79)$, $(\gamma_1, \gamma_2) = (0.1, 2)$. Dies ist die Parameterkombination, welche sich unter der Annahme lokaler Alternativen (3.2) für die Parameter $\theta_0 = (4.5, 1.81, 0.79)^T$, $\gamma_0 = (0, 1)^T$ und $\delta/\sqrt{150} = (0.1, 1)^T$ ergibt, welche unter Anderem zur Versuchsplanung genutzt wurde. Weitere Ergebnisse finden sich in Abschnitt 3.6.

In jedem Simulationsdurchlauf schätzen wir den Parameter $\mu = ED_{0.6}$ mittels Model-Averaging unter Verwendung der Beobachtungen $y_{11}^{(l)}, \dots, y_{kn_k}^{(l)}$ aus (3.29) für die verschiedenen Versuchspläne ξ_A^* , ξ_1 und ξ_2 . Den Schätzer der $ED_{0.6}$ in den einzelnen Simulationsdurchläufen bezeichnen wir im Folgenden mit $\hat{\mu}_{\text{mav}}^{(l)}$. Mit den so resultie-

renden 1000 Schätzungen der $ED_{0.6}$ berechnet sich dann der mittlere quadratische Fehler über

$$\text{MSE}(\xi) = \frac{1}{1000} \sum_{l=1}^{1000} \left(\hat{\mu}_{\text{mav}}^{(l)} - \mu_{\text{true}} \right)^2,$$

wobei μ_{true} die $ED_{0.6}$ im wahren sigmoidalen Emax-Modell (3.29) ist unter Verwendung der Parameter $(\vartheta^T, \gamma^T) = (1.81, 0.79, 0.1, 2)$.

Der simulierte mittlere quadratische Fehler des Model-Averaging Schätzers mit festen Gewichten $w_s = 0.25$, $s = 1, \dots, 4$, für die verschiedenen Versuchspläne ξ_A^* , ξ_1 und ξ_2 ist in der linken Spalte von Tabelle 3.1 zu sehen. Die zweite Spalte zeigt den mittleren quadratischen Fehler des Model-Averaging Schätzers mit den glatten AIC-Gewichten aus (2.24) und die dritte Spalte präsentiert die entsprechenden Ergebnisse unter Verwendung der Schätzer der $ED_{0.6}$ in dem Modell, welches mittels AIC-Kriterium ausgewählt wurde. In der vierten Spalte finden sich die Ergebnisse für den Schätzer im größten Modell. Die fettgedruckten Zahlen in jeder Spalte weisen auf den jeweils kleinsten mittleren quadratischen Fehler innerhalb der drei Versuchspläne hin.

Es ist zu beobachten, dass die Model-Averaging Schätzer stets einen geringeren mittleren quadratischen Fehler liefern als die Schätzer in dem vom AIC gewählten Modell. Zum Beispiel ergibt sich bei Verwendung des Versuchsplans ξ_A^* für den Schätzer nach Modellwahl ein mittlerer quadratischer Fehler von 0.596, während dieser lediglich 0.355 beziehungsweise 0.508 beträgt, wenn wir den Model-Averaging Schätzer mit fester (gleicher) Gewichtung beziehungsweise mit glatter AIC-Gewichtung verwenden (siehe Zeile 1 in Tabelle 3.1). Auch die anderen beiden (nicht optimalen) Versuchspläne zeigen ähnliche Ergebnisse. Diese Erkenntnis, sowie auch weitere Ergebnisse in Abschnitt 3.6, stimmen mit Beobachtungen von Schorning et al. (2016), Aoki et al. (2017) und Buatois et al. (2018) überein und deuten an, dass Model-Averaging präzisere Schätzungen liefert als der Schätzer in einem einzigen per Modellwahl bestimmten Modell (gemessen am mittleren quadratischen Fehler). Mit den Ergebnissen aus Tabelle 3.1 stellen wir zudem fest, dass der Model-Averaging Schätzer mit fester Gewichtung wesentlich kleinere mittlere quadratische Fehler aufweist als der Schätzer mit datenbasierten Gewichten. Einen ähnlichen Effekt beobachten auch Wagner und Hlouskova (2015) im Kontext der *principal components augmented regression*. Eine mögliche Erklärung für diese Beobachtung ist, dass die Model-Averaging Schätzung mit datenbasierten Gewichten eine größere Variabilität beinhaltet als die Schätzung mit festen Gewichten. Bei einer zugleich kleinen Verzerrung kann dies zu einem geringeren mittleren quadratischen Fehler unter Verwendung fester Gewichte füh-

Versuchsplan	Schätzmethode			
	Feste Gew.	Glatte AIC-Gew.	Modellselektion	Größtes Modell
ξ_A^*	0.355	0.508	0.596	0.675
ξ_1	0.810	0.913	1.017	1.267
ξ_2	0.637	0.846	0.994	1.121

Tabelle 3.1: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.6}$ mit Gewichten $w_s = 0.25, s = 1, \dots, 4$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen gehören zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_A^* zur Model-Averaging Schätzung der $ED_{0.6}$ aus (3.26). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

ren. Durch eine geschickte Wahl der Modellgewichte besteht die Möglichkeit diesen Sachverhalt umzukehren. Zahlreiche Autoren untersuchen die optimale Wahl von Gewichten, sodass der Model-Averaging Schätzer einen möglichst kleinen mittleren quadratischen Fehler besitzt (siehe etwa Hjort und Claeskens (2003), Hansen (2007) und Liang et al. (2011)). Diese Gewichte hängen jeweils von den untersuchten Modellen sowie den Zielparametern ab und sind im Allgemeinen datenbasiert, sodass wir diesen Ansatz nicht weiter verfolgen. Die vierte Spalte in Tabelle 3.1 zeigt, dass der Schätzer der $ED_{0.6}$ im größten Modell die größten mittleren quadratischen Fehler besitzt. Auch wenn dieser Schätzer unverfälscht ist, so besitzt er eine größere Varianz als die Schätzer in den kleineren Modellen, wodurch die Benutzung dieses Schätzers zu einem größeren mittleren quadratischen Fehler führt.

Verglichen mit den gleichverteilten Versuchsplänen ξ_1 und ξ_2 generiert der optimale Versuchsplan aus (3.26) eine Reduzierung des mittleren quadratischen Fehlers um 56% beziehungsweise 44% für den Schätzer mit den festen Modellgewichten. Weiterhin verringert dieser Versuchsplan auch den mittleren quadratischen Fehler bei Gebrauch der glatten AIC-Gewichtung (um 44% beziehungsweise 40%), bei Schätzung der $ED_{0.6}$ im durch das AIC gewählten Modell (um 41% beziehungsweise 40%) und bei Schätzung der $ED_{0.6}$ im größten Modell (um 47% beziehungsweise 40%).

Als weiteres Beispiel betrachten wir nun den Model-Averaging Schätzer (2.22) der effektiven Dosis $ED_{0.6}$ für die vier Modelle aus Beispiel 3.1, wobei wir nun im Gegensatz zu zuvor eine ungleiche Gewichtung gebrauchen, nämlich

$$w_1 = 0.1, w_2 = 0.1, w_3 = 0.3 \text{ und } w_4 = 0.5.$$

Versuchsplan	Schätzmethode			
	Feste Gew.	Glatte AIC-Gew.	Modellselektion	Größtes Modell
ξ_B^*	0.476	0.502	0.582	0.562
ξ_1	0.915	0.900	1.014	1.135
ξ_2	0.869	0.949	1.067	1.103

Tabelle 3.2: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.6}$ mit Gewichten $w_1 = 0.1, w_2 = 0.1, w_3 = 0.3$ und $w_4 = 0.5$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen gehören zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_B^* zur Model-Averaging Schätzung der $ED_{0.6}$ aus (3.30). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Der Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der $ED_{0.6}$ ist dann gegeben durch

$$\xi_B^* = \left\{ \begin{array}{ccccc} 0 & 0.809 & 1.665 & 2.691 & 8 \\ 0.152 & 0.120 & 0.175 & 0.279 & 0.274 \end{array} \right\}. \quad (3.30)$$

Für diesen Versuchsplan ist die notwendige Bedingung in der rechten Grafik in Abbildung 3.2 dargestellt. Der Vergleich der Versuchspläne ξ_A^* und ξ_B^* aus (3.26) und (3.30) zeigt, dass die Trägerpunkte zwar ähnlich sind, bei den Gewichten jedoch deutliche Unterschiede auftreten.

Auch für diesen Versuchsplan führen wir eine Simulationsstudie analog zur vorherigen durch. Die entsprechenden Ergebnisse finden sich in Tabelle 3.2. Diese Tabelle zeigt ein ähnliches Bild wie Tabelle 3.1, jedoch ist dieses weniger stark ausgeprägt. Die Model-Averaging Schätzer weisen wesentlich bessere Ergebnisse auf als die Schätzer in dem durch das AIC gewählten Modell (Verbesserung zwischen 10% und 19% unter Verwendung der festen Gewichtung und zwischen 11% und 14% unter Verwendung der glatten AIC-Gewichtung). Weiterhin ist bei Gebrauch von festen Gewichten anstelle von glatten AIC-Gewichten für die Versuchspläne ξ_B^* und ξ_2 eine Verbesserung des mittleren quadratischen Fehlers zu beobachten. Diese tritt bei Verwendung von ξ_1 nicht auf. Anders als in Tabelle 3.1 ist in dieser Situation der Schätzer im größten Modell präziser als der Schätzer nach Modellselektion mittels des AIC bei Verwendung des Bayes-optimalen Versuchsplans ξ_B^* . Weiterhin liefern jedoch die Model-Averaging Schätzer kleinere mittlere quadratische Fehler als der Schätzer im größten Modell. Der Vergleich von Tabelle 3.1 mit Tabelle 3.2 zeigt, dass für alle betrachteten Versuchspläne der Model-Averaging Schätzer mit (festen)

unterschiedlichen Modellgewichten größere mittlere quadratische Fehler besitzt als der Model-Averaging Schätzer mit (festen) gleichen Modellgewichten.

Der Bayes-optimale Versuchsplan ξ_B^* zur Model-Averaging Schätzung der $ED_{0.6}$ bietet gegenüber den Versuchsplänen ξ_1 und ξ_2 eine Verbesserung des mittleren quadratischen Fehlers um 48% beziehungsweise 45% bei Nutzung des Model-Averaging Schätzers mit festen (ungleichen) Gewichten. Unter Verwendung der übrigen Schätzmethoden ergibt sich ebenfalls eine Verringerung des mittleren quadratischen Fehlers um 43% bis 50%.

In Abschnitt 3.6.1 präsentieren wir in Tabelle 3.7 und Tabelle 3.8 weitere Simulationsergebnisse, welche auf anderen Parameterkombinationen in dem Modell (3.29) beruhen. Die enthaltenen Ergebnisse zeigen ein sehr ähnliches Bild wie das oben beschriebene. In allen untersuchten Szenarien, unabhängig von Versuchsplan und Parameterkombination, liefern die Model-Averaging Schätzer kleinere simulierte mittlere quadratische Fehler als der Schätzer in dem durch das AIC gewählten Modell. Ebenso resultieren unabhängig von der angewendeten Schätzmethode bei Verwendung der beiden Bayes-optimalen Versuchspläne zur Model-Averaging Schätzung der $ED_{0.6}$ in beinahe allen Fällen präzisere Schätzer als bei Verwendung der gleichverteilten Versuchspläne.

3.3.2 Schätzung der AUC im logistischen Modell

In diesem Abschnitt betrachten wir den Fall, dass $f_{\text{wide}} = f_4$ die Dichte einer Normalverteilung mit Varianz σ^2 und Erwartungswertfunktion

$$\eta_4(x, \vartheta, \gamma) = \gamma_1 + \frac{\vartheta_1}{1 + \exp[(\vartheta_2 - x)/\gamma_2]}, \quad x \geq 0, \quad (3.31)$$

ist. Dieses logistische Modell wird ebenso wie das sigmoidale Emax-Modell zur Modellierung von Dosis-Wirkungs-Beziehungen genutzt, aber auch um das Wachstum von Populationen zu beschreiben (siehe beispielsweise Zwietering (1990)). Auch hier ist der Versuchsraum gegeben durch $\mathcal{X} = [0, 8]$, in dem $n = 150$ Experimente durchgeführt werden können. Das Ziel sei nun, die Fläche unter der Kurve (AUC) zu schätzen, welche in (3.7) definiert ist. Dabei wählen wir den Bereich \mathcal{C} identisch zum Versuchsraum. In der Terminologie von Dosis-Wirkungs-Studien lassen sich in (3.31) der Wert $\eta(0, \vartheta, \gamma)$ als Placebo-Effekt, ϑ_1 als asymptotischer maximaler Effekt des Medikamentes (relativ zum Placebo-Effekt) und $\vartheta_2 > 0$ als die Dosis, bei der die Hälfte des maximalen Effekts erreicht wird, interpretieren. Der Parameter γ_2 bestimmt die Steigung der Erwartungswertfunktion η_4 . Wir gehen im Folgenden davon aus, dass die beiden für Medikamentenstudien zentralen Parameter ϑ_1 und ϑ_2

sowie die Varianz σ^2 in jedem Kandidatenmodell enthalten sind und fassen diese zusammen in $\theta = (\sigma^2, \vartheta_1, \vartheta_2)^T$. Die Komponenten des Parametervektors $\gamma = (\gamma_1, \gamma_2)^T$ dagegen können auf die entsprechenden Bestandteile von $\gamma_0 = (0, 1)^T$ fixiert werden, sodass $r = 4$ Kandidatenmodelle in der Menge \mathcal{S} enthalten sind. Dies sind jeweils Regressionsmodelle der Form (2.1), welche durch unterschiedliche Erwartungswertfunktionen charakterisiert sind, nämlich

$$\eta_1(x, \vartheta) = \frac{\vartheta_1}{1 + \exp[(\vartheta_2 - x)]}, \quad (3.32)$$

$$\eta_2(x, \vartheta, (0, \gamma_2)) = \frac{\vartheta_1}{1 + \exp[(\vartheta_2 - x)/\gamma_2]}, \quad (3.33)$$

$$\eta_3(x, \vartheta, (\gamma_1, 1)) = \gamma_1 + \frac{\vartheta_1}{1 + \exp[(\vartheta_2 - x)]} \quad (3.34)$$

und η_4 aus (3.31). Da die Parameter γ_1 und ϑ_1 linear in die Modelle eingehen, benötigen wir lediglich für die Parameter γ_2 und ϑ_2 A-priori Verteilungen zur Bestimmung Bayes-optimaler Versuchspläne zur Model-Averaging Schätzung der Fläche unter der Kurve. Diese seien unabhängige Gleichverteilungen auf den Mengen $\{3, 4, 5\}$ beziehungsweise $\{5/6, 1, 7/6\}$. Für die Varianz wählen wir eine Einpunktverteilung in $\sigma_0^2 = 4.5$ und der Parameter δ ist so bestimmt, dass $\delta^T/\sqrt{150} = (0.015, -1/6)$ ist. Mit diesen Spezifikationen ist der numerisch bestimmte Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der AUC mit gleicher Gewichtung

$$w_s = 0.25, s = 1, \dots, 4,$$

gegeben durch

$$\xi_C^* = \left\{ \begin{array}{ccccc} 0 & 2.585 & 4.332 & 5.419 & 8 \\ 0.094 & 0.258 & 0.239 & 0.204 & 0.206 \end{array} \right\}. \quad (3.35)$$

Dieser Versuchsplan erfüllt die notwendige Bedingung aus Satz 3.7 (hier nicht gezeigt).

Auch für das logistische Modell untersuchen wir die Eigenschaften verschiedener Versuchspläne mittels einer Simulationsstudie, in der die Daten aus dem Modell

$$y_{ij}^{(l)} = \gamma_1 + \frac{\vartheta_1}{1 + \exp[(\vartheta_2 - x)/\gamma_2]} + \sigma \varepsilon_{ij}^{(l)}, i = 1, \dots, k, j = 1, \dots, n_i, \quad (3.36)$$

generiert werden, wobei $\varepsilon_{ij}^{(l)}$ unabhängige standardnormalverteilte Zufallsvariablen sind und wir $n = \sum_{i=1}^k n_i = 150$ Experimente durchführen können.

Im Weiteren beschränken wir uns darauf, die Ergebnisse für die Parameterkombination $\vartheta^T = (\vartheta_1, \vartheta_2) = (-1.73, 4)$, $\gamma^T = (0.015, 0.833)$ und $\sigma^2 = 4.5$ zu diskutieren.

Versuchsplan	Schätzmethode			
	Feste Gew.	Glatte AIC-Gew.	Modellselektion	Größtes Modell
ξ_C^*	1.659	1.880	2.074	2.071
ξ_1	1.961	2.080	2.196	2.059
ξ_2	1.687	1.763	1.838	1.847

Tabelle 3.3: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der Fläche unter der Kurve mit Gewichten $w_s = 0.25, s = 1, \dots, 4$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen gehören zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_C^* zur Model-Averaging Schätzung der AUC aus (3.35). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Dies entspricht unter der Annahme der lokalen Alternativen (3.2) der Situation, dass $\theta_0^T = (4.5, -1.73, 4)$, $\gamma_0^T = (0, 1)$ und $\delta^T/\sqrt{150} = (0.015, -1/6)$ sind. Weitere Resultate mit anderen Parameterkombinationen zeigen ein ähnliches Bild und sind in Abschnitt 3.6 zu finden.

Der mittlere quadratische Fehler des Model-Averaging Schätzers mit gleichen Modellgewichten $w_s = 0.25, s = 1, \dots, 4$, für die verschiedenen Versuchspläne ist in der ersten Spalte von Tabelle 3.3 abgetragen, während die zweite und dritte Spalte die entsprechenden Ergebnisse für den Model-Averaging Schätzer basierend auf der glatten AIC-Gewichtung beziehungsweise den Schätzer nach Modellselektion mittels AIC zeigen. In der vierten Spalte sind die mittleren quadratischen Fehler des Schätzers der $ED_{0.6}$ im größten Modell zu finden.

Wie in Abschnitt 3.3.1 ist auch in Tabelle 3.3 eine präzisere Schätzung der Fläche unter der Kurve unter Verwendung eines der Model-Averaging Schätzverfahrens anstelle des Schätzers nach Modellselektion zu beobachten. In Zahlen ausgedrückt verringert sich bei Gebrauch von festen Modellgewichten der mittlere quadratische Fehler zwischen 8% und 20% (abhängig vom betrachteten Versuchsplan), während wir durch Nutzung von Model-Averaging mit glatten AIC-Gewichten eine Verbesserung zwischen 4% und 9% erzielen können. Des Weiteren liefert der Model-Averaging Schätzer bessere Resultate bei Verwendung von festen (gleichen) Gewichten anstelle von glatten AIC-Gewichten. Zudem sind die mittleren quadratischen Fehler des Model-Averaging Schätzers mit festen Gewichten kleiner als die mittleren quadratischen Fehler des Schätzers der $ED_{0.6}$ im größten Modell.

Betrachten wir den Model-Averaging Schätzer mit festen (gleichen) Gewichten, so erkennen wir, dass der Bayes-optimale Versuchsplan ξ_C^* zur Model-Averaging Schätzung der Fläche unter der Kurve den gleichverteilten Versuchsplan ξ_1 um 15% verbessert (im Sinne des mittleren quadratischen Fehlers), während die Verbesserung gegenüber dem Versuchsplan ξ_2 lediglich 2% beträgt. Nutzen wir dagegen datenbasierte Gewichte, also glatte AIC-Gewichte oder Modellselektion, oder den Schätzer im größten Modell, so ist der mittlere quadratische Fehler bei Verwendung des gleichverteilten Versuchsplans ξ_2 am geringsten. Dies lässt sich durch die Tatsache, dass ξ_C^* nicht für eine derartige Schätzung konstruiert ist, erklären. Folglich liefern die anhand des Kriteriums aus Definition 3.6 bestimmten Versuchspläne in vielen Fällen zwar gute Ergebnisse, wir können jedoch im Allgemeinen nicht davon ausgehen, dass der resultierende optimale Versuchsplan auch nah am optimalen Versuchsplan zur Model-Averaging Schätzung mit glatten AIC-Gewichten beziehungsweise zur Schätzung in dem durch das AIC gewählten Modell ist. Ungeachtet dessen generiert der Model-Averaging Schätzer basierend auf festen Gewichten gepaart mit dem Bayes-optimale Versuchsplan den kleinsten mittleren quadratischen Fehler in Tabelle 3.3. Nun betrachten wir einen Model-Averaging Schätzer mit (ungleichen) Modellgewichten

$$w_1 = 0.1, w_2 = 0.1, w_3 = 0.1 \text{ und } w_4 = 0.7$$

für die jeweiligen Modelle (3.32), (3.33), (3.34) und (3.31). Der entsprechende Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der Fläche unter der Kurve ist dann

$$\xi_D^* = \left\{ \begin{array}{ccccc} 0 & 2.418 & 4.259 & 5.777 & 8 \\ 0.122 & 0.284 & 0.197 & 0.253 & 0.145 \end{array} \right\}. \quad (3.37)$$

Mit analogen Simulationen wie zuvor erhalten wir für die verschiedenen Versuchspläne die mittleren quadratischen Fehler in Tabelle 3.4. Diese Ergebnisse bestätigen die Beobachtungen aus Abschnitt 3.3.1, nämlich, dass der Model-Averaging Schätzer wesentlich präziser ist als der Schätzer nach Modellselektion. Anders als in Abschnitt 3.3.1 liefert hier aber der Model-Averaging Schätzer mit glatten AIC-Gewichten leicht bessere Ergebnisse als der Model-Averaging Schätzer mit festen Gewichten (es ist ein Anstieg von etwa 2% zu beobachten bei Verwendung der festen Gewichte). Während in Tabelle 3.3 der Schätzer der Fläche unter der Kurve im größten Modell teils bessere Ergebnisse aufweist als der Schätzer im mittels AIC gewählten Modell, zeigt Tabelle 3.4 für den Schätzer im größten Modell die höchsten mittleren quadratischen Fehler aller Schätzmethoden, unabhängig vom betrachteten Versuchsplan.

Versuchsplan	Schätzmethode			
	Feste Gew.	Glatte AIC-Gew.	Modellselektion	Größtes Modell
ξ_D^*	1.764	1.723	1.835	1.978
ξ_1	2.059	2.041	2.129	2.185
ξ_2	1.841	1.801	1.883	1.886

Tabelle 3.4: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der Fläche unter der Kurve mit Gewichten $w_1 = w_2 = w_3 = 0.1$ und $w_4 = 0.7$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen gehören zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_D^* zur Model-Averaging Schätzung der AUC aus (3.37). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Für alle Model-Averaging Verfahren (inklusive des Schätzers nach Modellselektion) resultiert die Verwendung des Versuchsplans ξ_D^* aus (3.37) in kleineren mittleren quadratischen Fehlern als die Verwendung der Versuchspläne ξ_1 und ξ_2 .

Weitere Simulationsergebnisse unter Verwendung anderer Parameterkombinationen sind in Abschnitt 3.6.2 in Tabelle 3.9 (Model-Averaging Schätzer mit gleichen Modellgewichten) und Tabelle 3.10 (Model-Averaging Schätzer mit unterschiedlichen Modellgewichten) zu finden und lassen ähnliche Schlüsse wie soeben beschrieben zu. Beispielsweise sind unabhängig vom Versuchsplan Model-Averaging Schätzer stets besser als der Schätzer in einem einzelnen, durch das AIC gewählten Modell. In den meisten Situationen bewirkt die Verwendung der Bayes-optimalen Versuchspläne zur Model-Averaging Schätzung der Fläche unter der Kurve verglichen mit den beiden gleichverteilten Versuchsplänen eine substanzielle Verbesserung des mittleren quadratischen Fehlers, selbst bei Gebrauch des Model-Averaging Schätzers mit glatten AIC-Gewichten oder des Schätzers nach Modellwahl. Einzelheiten hierzu finden sich in Abschnitt 3.6.2.

Bemerkung. In den Beispielen in Abschnitt 3.3.1 und Abschnitt 3.3.2 haben wir jeweils Simulationsstudien präsentiert, in denen wir Model-Averaging Schätzer mit zufälligen Gewichten auf Grundlage des AIC-Kriteriums sowie Schätzer in dem Modell, das mittels AIC ausgewählt wurde, verwendet haben. Wir haben die gleichen Simulationen wie zuvor zusätzlich unter Verwendung von BIC-Gewichten durchgeführt, indem wir in (2.23) und (2.24) jeweils das BIC-Kriterium aus (2.21) anstelle des AIC-Kriteriums eingesetzt haben (siehe auch Claeskens und Hjort (2008, Kapi-

tel 7)). Die Ergebnisse sind den hier gezeigten Ergebnissen jedoch sehr ähnlich und wir geben sie daher nicht gesondert an. Insbesondere sind keine wesentlichen Unterschiede zwischen den mittleren quadratischen Fehlern der Model-Averaging Schätzer unter Verwendung der glatten BIC-Gewichtung und der Schätzer nach Modellselektion mit dem BIC im Vergleich zu denen, die auf dem AIC basieren, zu beobachten.

3.4 Ein adaptiver Ansatz

In Abschnitt 3.3 zeigen die Simulationsergebnisse, dass Model-Averaging Schätzer basierend auf zufälligen Gewichten oftmals unpräziser sind als diejenigen, die feste Modellgewichte nutzen. In der Praxis jedoch ist es üblich zufällige Gewichte, wie die glatte AIC-Gewichtung oder die Modellselektion zu verwenden (siehe beispielsweise Verrier et al. (2014)). Auch wenn die Bayes-optimalen Versuchspläne zur Schätzung des Parameters μ in Abschnitt 3.3 oftmals auch die mittleren quadratischen Fehler der Model-Averaging Schätzer basierend auf datenbasierten Modellgewichten verringern, deuten die Ergebnisse aus den Abschnitten 3.3.1, 3.3.2, 3.6.1 und 3.6.2 darauf hin, dass wir die Schätzgenauigkeit zusätzlich verbessern können, indem wir alternative Verfahren entwickeln, welche die Schätzung mit datenbasierten Gewichten weiter verbessern. Dies soll das Ziel dieses Abschnitts sein.

Wie in Abschnitt 3.1 diskutiert, ist der asymptotische mittlere quadratische Fehler des Model-Averaging Schätzers mit datenbasierten Gewichten nicht in geschlossener Form, wie etwa in (3.20), darstellbar, sodass hier auf Simulationen oder numerische Integration zurückgegriffen werden müsste, um ihn zu bestimmen. Damit ist die direkte Konstruktion eines Versuchsplanungskriteriums für Model-Averaging Schätzer mit zufälligen Gewichten nicht ohne Weiteres möglich. Betrachten wir aber beispielsweise die Situation aus Abschnitt 3.3.2, so zeigt ein Vergleich der mittleren quadratischen Fehler der Model-Averaging Schätzer mit zufälligen Gewichten, dass der Versuchsplan ξ_D^* (konstruiert für ungleiche, feste Modellgewichte) präzisere Schätzer ermöglicht als der Versuchsplan ξ_C^* (konstruiert für gleiche, feste Modellgewichte) (vergleiche Tabelle 3.3 und Tabelle 3.4, Verbesserung um 8.4% beziehungsweise 11.5%). Daher scheint es möglich, über eine geschickte Wahl der Modellgewichte im Versuchsplanungsschritt den mittleren quadratischen Fehler des Model-Averaging Schätzers mit zufälligen Gewichten zu verringern. Aus diesem Grund schlagen wir in diesem Abschnitt eine adaptive Vorgehensweise vor. Hierbei beschränken wir uns zur Illustration auf einen Ansatz basierend auf zwei Schritten mit dem Ziel der Model-

Averaging Schätzung mit glatten AIC-Gewichten (2.24). Die Erweiterung auf mehr Schritte oder andere Gewichte erfolgt analog zum beschriebenen Vorgehen.

Unser adaptiver Ansatz basiert auf einer Aufteilung der n möglichen Beobachtungen in zwei Stichproben der Größe $n^{(1)}$ und $n^{(2)}$. Die ersten $n^{(1)}$ Experimente führen wir gemäß eines Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung des Parameters μ mit festen Gewichten durch. Dabei nutzen wir im Folgenden jeweils gleiche Modellgewichte, die Verwendung anderer Gewichte ist aber ebenso möglich. Wir bezeichnen die erste Stichprobe mit $Y_{n^{(1)}}^{(1)}$ und berechnen basierend auf dieser die glatten AIC-Gewichte $w_s^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$ ($s = 1, \dots, r$). Die übrigen $n^{(2)}$ Experimente führen wir entsprechend eines Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung des Parameters μ mit Gewichten $w_1^{\text{smAIC}}(Y_{n^{(1)}}^{(1)}), \dots, w_r^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$ durch. Es ist zu beachten, dass der Model-Averaging Schätzer bedingt auf die erste Stichprobe $Y_{n^{(1)}}^{(1)}$ asymptotisch normalverteilt ist (für $n^{(2)} \rightarrow \infty$), wodurch wir das Optimalitätskriterium aus Definition 3.6 nutzen können. In den Simulationen nutzen wir später die gesamte Stichprobe mit $n = n^{(1)} + n^{(2)}$ Beobachtungen, um die mittleren quadratischen Fehler des Model-Averaging Schätzers des Parameters μ beziehungsweise des Schätzers nach Modellselektion mit dem AIC zu bestimmen. Wir betrachten im Folgenden die beiden Beispiele aus Abschnitt 3.3.

3.4.1 Schätzung der ED_α im sigmoidalen Emax-Modell

Wir betrachten das sigmoidale Emax-Modell (3.3) mit den gleichen Parametern wie in Abschnitt 3.3.1, wobei die Erwartungswertfunktionen der weiteren Kandidatenmodelle in (3.4) beziehungsweise (3.5) zu finden sind. Wir teilen die Stichprobe von $n = 150$ Beobachtungen auf in zwei Gruppen der Größe $n^{(1)} = 50$ und $n^{(2)} = 100$ und nutzen den Bayes-optimalen Versuchsplan zur Model-Averaging Schätzung der $\text{ED}_{0.6}$ mit festen Gewichten ($w_s = 0.25$, $s = 1, \dots, 4$) um die ersten $n^{(1)} = 50$ Versuche durchzuführen. Darauf basierend berechnen wir die glatten AIC-Gewichte $w_s^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$, $s = 1, \dots, 4$, und erheben die übrigen $n^{(2)} = 100$ Beobachtungen gemäß des Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung der $\text{ED}_{0.6}$ mit den Gewichten $w_1^{\text{smAIC}}(Y_{n^{(1)}}^{(1)}), \dots, w_4^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$. Hierbei ist die Aufteilung der Stichprobe zunächst willkürlich. In Kapitel 4 findet sich ein Ansatz, wie die Aufteilung der Stichprobe erfolgen kann, wobei hier Simulationen verwendet werden.

Mit den so konstruierten Versuchsplänen führen wir eine Simulationsstudie analog zu Abschnitt 3.3.1 beziehungsweise Abschnitt 3.6.1 durch. Wir generieren also Daten gemäß (3.29) für verschiedene Parameterkombinationen (ϑ, γ) , während wir die Varianz $\sigma^2 = 4.5$ konstant halten. Die Ergebnisse finden sich in Tabelle 3.5. Diese

Parameter (ϑ, γ)	Adaptives Vorgehen	
	Glatte AIC-Gew.	Modellselektion
(1.81,0.79,0,1)	1.241	1.579
(1.81,0.79,0.1,1)	1.091	1.317
(1.81,0.79,0,2)	0.584	0.683
(1.81,0.79,0.1,2)	0.604	0.686
(1.81,1.79,0,2)	0.838	1.024
(1.81,1.79,0.1,2)	0.974	1.159
(1.81,1.79,0,3)	0.610	0.680
(1.81,1.79,0.1,3)	0.586	0.657

Tabelle 3.5: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.6}$ im sigmoidalen Emax-Modell unter Verwendung des adaptiven Vorgehens mit glatten AIC-Gewichten $w_1^{\text{smAIC}}(Y_n), \dots, w_4^{\text{smAIC}}(Y_n)$ (erste Spalte) und des Schätzers nach Modellselektion (zweite Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen.

Tabelle zeigt die Ergebnisse für zwei unterschiedliche adaptive Vorgehensweisen. Bei dem oben beschriebenen Vorgehen haben wir insgesamt $n = n^{(1)} + n^{(2)}$ Beobachtungen vorliegen und bestimmen Schätzer $\hat{\mu}_s$ für die $ED_{0.6}$ gemäß (3.6) in jedem Kandidatenmodell ($s = 1, \dots, 4$). Als Gewichte für den Model-Averaging Schätzer (2.22) wählen wir nun entweder die glatten AIC-Gewichte basierend auf der Gesamtstichprobe $w_1^{\text{smAIC}}(Y_n), \dots, w_4^{\text{smAIC}}(Y_n)$ oder die Gewichte $w_1^{\text{AIC}}(Y_n), \dots, w_4^{\text{AIC}}(Y_n)$, mit denen sich der Schätzer nach Modellselektion mit dem AIC ergibt. Die aus der Simulation resultierenden mittleren quadratischen Fehler des Model-Averaging Schätzers mit glatten AIC-Gewichten finden sich in der ersten Spalte von Tabelle 3.5, während die Ergebnisse für den Schätzer nach Modellselektion in der zweiten Spalte zu finden sind. Das Ziel der adaptiven Vorgehensweise ist, die Model-Averaging Schätzung mit datenbasierten Gewichten im Vergleich zu den Versuchsplänen ξ_A^* und ξ_B^* aus Abschnitt 3.3.1 zu verbessern. Daher sind die Ergebnisse mit denen aus Tabelle 3.7 und Tabelle 3.8 zu vergleichen.

Zunächst zeigt Tabelle 3.5, dass Model Averaging Schätzer mit glatten AIC-Gewichten präziser sind als die Schätzer nach Modellselektion mit dem AIC. Dies haben wir bereits in Abschnitt 3.3.1 beobachtet. Es zeigt sich jedoch auch, dass das adaptive Vorgehen in den meisten Szenarien (unabhängig von der Schätzmethode) größere mittlere quadratische Fehler aufweist als die optimalen Versuchspläne ξ_A^* aus (3.26) und ξ_B^* aus (3.30). Betrachten wir beispielsweise die Situation aus Tabelle 3.1 be-

ziehungsweise Tabelle 3.2, so beträgt der mittlere quadratische Fehler des Model-Averaging Schätzers mit glatten AIC-Gewichten 0.508 beziehungsweise 0.502 unter Verwendung der Versuchspläne ξ_A^* beziehungsweise ξ_B^* . Weiterhin ist der mittlere quadratische Fehler mit diesen Versuchsplänen 0.596 und 0.582 bei Gebrauch des Schätzers nach Modellselektion. Für die entsprechende Parameterkombination weist das adaptive Vorgehen dagegen einen mittleren quadratischen Fehler von 0.604 beziehungsweise 0.686 unter Verwendung des Model-Averaging Schätzers beziehungsweise des Schätzers nach Modellselektion auf. Diese Erhöhung des mittleren quadratischen Fehlers ist darauf zurückzuführen, dass der Schätzer durch die Verwendung zufälliger Gewichte eine zusätzliche Variabilität erhält, sodass oftmals die Schätzung mit festen Gewichten präziser ist als die Schätzung mit zufälligen Gewichten (siehe Abschnitt 3.3.1). In der adaptiven Vorgehensweise werden nun jeweils zweimal datenbasierte Gewichte bestimmt, welche jeweils zusätzliche Varianz mit sich bringen. Insgesamt zeigt ein Vergleich von Tabelle 3.5 mit Tabelle 3.7 und Tabelle 3.8, dass die adaptive Vorgehensweise in den meisten untersuchten Szenarien unpräziser ist als die nicht-adaptive Vorgehensweise. Dennoch resultieren unter Verwendung dieser Methode stets kleinere mittlere quadratische Fehler als unter Verwendung von ξ_1 aus (3.27) und ξ_2 aus (3.28).

3.4.2 Schätzung der AUC im logistischen Modell

In diesem Abschnitt untersuchen wir adaptive Versuchspläne zur Model-Averaging Schätzung der Fläche unter der Kurve im logistischen Modell (3.31). Die Erwartungswertfunktionen der übrigen Kandidatenmodelle sind in (3.32), (3.33) und (3.34) zu finden. Das Vorgehen ist dabei das Gleiche wie jenes zu Beginn von Abschnitt 3.4.1 beschriebene. Auch hier wählen wir $n^{(1)} = 50$ und $n^{(2)} = 100$.

Die Durchführung einer Simulationsstudie analog zu Abschnitt 3.3.2 beziehungsweise Abschnitt 3.6.2 führt zu den Ergebnissen aus Tabelle 3.6. Auch hier ist in der ersten Spalte der mittlere quadratische Fehler des adaptiven Vorgehens für den Model-Averaging Schätzer mit glatten AIC-Gewichten zu sehen, während die zweite Spalte die entsprechenden Ergebnisse für den Schätzer nach Modellselektion präsentiert. Diese sind mit den mittleren quadratischen Fehlern der Bayes-optimalen Versuchspläne zur Model-Averaging Schätzung der Fläche unter der Kurve ξ_C^* aus Tabelle 3.9 und ξ_D^* aus Tabelle 3.10 zu vergleichen.

Auch in Tabelle 3.6 zeigt sich, dass der Schätzer nach Modellselektion stets größere mittlere quadratische Fehler besitzt als der Model-Averaging Schätzer. Wir betrachten zunächst die Simulationsergebnisse für die Situation aus Abschnitt 3.3.2. Für

Parameter (ϑ, γ)	Adaptives Vorgehen	
	Glatte AIC-Gew.	Modellselektion
(-1.73,4,0,1)	1.855	1.907
(-1.73,4,0.015,1)	1.675	1.719
(-1.73,4,0,0.833)	1.793	1.795
(-1.73,4,0.015,0.833)	1.646	1.657
(-1.73,5,0,0.833)	1.779	1.858
(-1.73,5,0.015,0.833)	1.608	1.680
(-1.73,5,0,0.667)	1.653	1.794
(-1.73,5,0.015,0.667)	1.594	1.679

Tabelle 3.6: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der AUC im logistischen Modell unter Verwendung des adaptiven Vorgehens mit glatten AIC-Gewichten $w_1^{\text{smAIC}}(Y_n), \dots, w_4^{\text{smAIC}}(Y_n)$ (erste Spalte) und des Schätzers nach Modellselektion (zweite Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen.

diese Parameterkombination besitzt das adaptive Vorgehen einen mittleren quadratischen Fehler von 1.646 beziehungsweise 1.657 unter Verwendung des Model-Averaging Schätzers beziehungsweise des Schätzers nach Modellselektion. Diese sind damit für den Model-Averaging Schätzer um 12.4% und 4.5% sowie für den Schätzer nach Modellselektion um 20.1% und 9.7% geringer als die entsprechenden Resultate der beiden Bayes-optimalen Versuchspläne ξ_C^* aus (3.35) und ξ_D^* aus (3.37) (siehe Tabelle 3.3 und Tabelle 3.4). Zudem zeigt Tabelle 3.3, dass der gleichverteilte Versuchsplan ξ_2 in den Simulationen kleinere mittlere quadratische Fehler erzielt als der Bayes-optimale Versuchsplan ξ_C^* bei Verwendung des Model-Averaging Schätzers mit glatten AIC-Gewichten oder des Schätzers nach Modellselektion. Das adaptive Vorgehen hingegen besitzt kleinere mittlere quadratische Fehler als der Versuchsplan ξ_2 . Insgesamt zeigt ein Vergleich mit den übrigen Szenarien aus Tabelle 3.9 und Tabelle 3.10, dass das adaptive Vorgehen für die Mehrheit der betrachteten Parameterkombinationen geringere mittlere quadratische Fehler hat als die Bayes-optimalen Versuchspläne. Darüber hinaus zeigt sich, dass dieses Vorgehen stets besser ist als die Verwendung der gleichverteilten Versuchspläne ξ_1 und ξ_2 , insbesondere auch in den Situationen, in denen die Bayes-optimalen Versuchspläne ξ_C^* und ξ_D^* schlechtere Resultate liefern als der Versuchsplan ξ_2 .

3.5 Beweise

Satz 3.3. Der Beweis der Aussage findet sich in Hjort und Claeskens (2003, Lemma 3.3), welche jedoch eine andere Notation verwenden. Daher zeigen wir im Folgenden die Gleichheit zu der von uns verwendeten Notation, wozu wir zunächst ihre Notation einführen müssen. Sie partitionieren zunächst die Informationsmatrix im größten Modell entsprechend der Parameter (θ, γ) , wir haben also

$$J(\xi, \theta_0, \gamma_0) = \begin{pmatrix} J_{00}(\xi, \theta_0, \gamma_0) & J_{01}(\xi, \theta_0, \gamma_0) \\ J_{10}(\xi, \theta_0, \gamma_0) & J_{11}(\xi, \theta_0, \gamma_0) \end{pmatrix}.$$

Weiterhin benötigen wir die Inverse der Informationsmatrix, welche wir ebenfalls partitionieren:

$$J(\xi, \theta, \gamma)^{-1} = \begin{pmatrix} J^{00}(\xi, \theta_0, \gamma_0) & J^{01}(\xi, \theta_0, \gamma_0) \\ J^{10}(\xi, \theta_0, \gamma_0) & J^{11}(\xi, \theta_0, \gamma_0) \end{pmatrix}.$$

Es ist zu beachten, dass für die Informationsmatrix aus (3.9) im Modell s gilt:

$$J_s(\xi, \theta_0, \gamma_{0,S_s}) = P_s J(\xi, \theta_0, \gamma_0) P_s^T, \quad (3.38)$$

mit P_s aus (3.15). Sofern es zu keinen Verwechslungen kommen kann, übergehen wir im Folgenden die Abhängigkeiten der Matrizen und Vektoren von den Parametern und dem Versuchsplan, wir schreiben also beispielsweise J_s anstelle von $J_s(\xi, \theta_0, \gamma_{0,S_s})$.

Unter Verwendung von Rechenregeln für Inversen von partitionierten Matrizen (siehe beispielsweise Lütkepohl (1996, Kapitel 3.5.3)) erhalten wir

$$K := J^{11} = (J_{11} - J_{10}J_{00}^{-1}J_{01})^{-1}.$$

Desweiteren führen Hjort und Claeskens (2003) die Matrizen K_s und K_s^0 ein, welche gegeben sind durch

$$K_s = (\pi_{S_s} K^{-1} \pi_{S_s}^T)^{-1}; \quad K_s^0 = \pi_{S_s}^T K_s \pi_{S_s}.$$

Abschließend schreiben sie

$$\omega = J_{10} J_{00}^{-1} \frac{\partial \mu}{\partial \theta} - \frac{\partial \mu}{\partial \gamma}. \quad (3.39)$$

Mit dieser Notation erhalten sie

$$\sqrt{n} \left(\hat{\mu}_s - \mu(\theta_0, \gamma_0 + \frac{\delta}{\sqrt{n}}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(\omega^T (I_{q \times q} - K_s^0 K^{-1}) \delta, \left(\frac{\partial \mu}{\partial \theta} \right)^T J_{\emptyset}^{-1} \left(\frac{\partial \mu}{\partial \theta} \right) + \omega^T K_s^0 \omega), \quad (3.40)$$

wobei sowohl in (3.39) als auch in (3.40) die Ableitungen im Parameter (θ_0, γ_0) auszuwerten sind und J_\emptyset die Informationsmatrix im kleinsten Modell f_{narrow} ist (Hjort und Claeskens, 2003, Lemma 3.3).

Wir zeigen zunächst die Gleichheit des Erwartungswertes aus (3.40) mit dem Erwartungswert (3.12). Mit L_s aus (3.14) ergibt sich:

$$\begin{aligned}
\omega^T (I_{q \times q} - K_s^0 K^{-1}) \delta &= c^T \begin{pmatrix} J_{00}(\xi)^{-1} J_{01}(\xi) \\ -I_{q \times q} \end{pmatrix} (I_{q \times q} - K_s^0 K^{-1}) \delta \\
&= c^T \left\{ \begin{pmatrix} J_{00}^{-1} J_{01} (I_{q \times q} - K_s^0 (J_{11} - J_{10} J_{00}^{-1} J_{01})) \\ K_s^0 (J_{11} - J_{10} J_{00}^{-1} J_{01}) \end{pmatrix} - \begin{pmatrix} 0 \\ I_{q \times q} \end{pmatrix} \right\} \delta \\
&= c^T \left\{ \begin{pmatrix} J_{00}^{-1} + J_{00}^{-1} J_{01} K_s^0 J_{10} J_{00}^{-1} & -J_{00}^{-1} J_{01} K_s^0 \\ -K_s^0 J_{10} J_{00}^{-1} & K_s^0 \end{pmatrix} \begin{pmatrix} J_{01} \\ J_{11} \end{pmatrix} - \begin{pmatrix} 0 \\ I_{q \times q} \end{pmatrix} \right\} \delta \\
&= c^T \left\{ \begin{pmatrix} J_{00}^{-1} + J_{00}^{-1} J_{01} K_s^0 J_{10} J_{00}^{-1} & -J_{00}^{-1} J_{01} \pi_s^T K_s \pi_s \\ -\pi_s^T K_s \pi_s J_{10} J_{00}^{-1} & \pi_s^T K_s \pi_s \end{pmatrix} \begin{pmatrix} J_{01} \\ J_{11} \end{pmatrix} - \begin{pmatrix} 0 \\ I_{q \times q} \end{pmatrix} \right\} \delta \\
&= c^T L_s(\xi, \vartheta_0) \delta.
\end{aligned}$$

Wir betrachten nun den Varianzterm. Hierzu ist folgende Gleichheit zu beachten:

$$\left(\frac{\partial \mu}{\partial \theta} \right)^T J_\emptyset^{-1} \left(\frac{\partial \mu}{\partial \theta} \right) = c_s^T \begin{pmatrix} J_{00}^{-1} & 0_{p \times |S_s|} \\ 0_{|S_s| \times p} & 0_{|S_s| \times |S_s|} \end{pmatrix} c_s, s = 1, \dots, r.$$

Damit erhalten wir

$$\begin{aligned}
\omega^T K_s^0 \omega &= c^T \begin{pmatrix} J_{00}^{-1} J_{01} \\ -I_{q \times q} \end{pmatrix} \pi_s^T K_s \pi_s \begin{pmatrix} J_{10} J_{00}^{-1} & -I_{q \times q} \end{pmatrix} c \\
&= c^T P_S^T \begin{pmatrix} J_{00}^{-1} J_{01} \pi_s^T \\ -I_{|S_s| \times |S_s|} \end{pmatrix} K_s \begin{pmatrix} \pi_s J_{10} J_{00}^{-1} & -I_{|S_s| \times |S_s|} \end{pmatrix} P_S c \\
&= c_s^T \begin{pmatrix} J_{00}^{-1} J_{01} \pi_s^T K_s \pi_s J_{10} J_{00}^{-1} & -J_{00}^{-1} J_{01} \pi_s^T K_s \\ -K_s \pi_s J_{10} J_{00}^{-1} & K_s \end{pmatrix} c_s.
\end{aligned}$$

Hiermit ergibt sich mit entsprechenden Formeln für die Inverse einer partitionierten Matrix (Lütkepohl, 1996, Kapitel 3.5.3) die abschließende Gleichheit

$$\left(\frac{\partial \mu}{\partial \theta} \right)^T J_\emptyset^{-1} \left(\frac{\partial \mu}{\partial \theta} \right) + \omega^T K_s^0 \omega = c_s^T J_S^{-1} c_s.$$

□

Satz 3.4. Auch in diesem Beweis schreiben wir die Abhängigkeiten der einzelnen Komponenten vom Versuchsplan ξ und den Parametern (θ_0, γ_0) nicht explizit auf.

Der Beweis vereinfacht sich substantziell, indem wir eine andere Notation für die asymptotische Verteilung in Satz 3.3 verwenden. Dazu definieren wir Zufallsvariablen U, V für die mit der Informationsmatrix J im größten Modell f_{wide} gilt:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}(0, J).$$

Darauf aufbauend gilt in einem Modell s ($s = 1, \dots, r$), dass

$$\begin{pmatrix} U \\ \pi_{S_s} V \end{pmatrix} \sim \mathcal{N}(0, P_s J P_s^T),$$

mit P_s aus (3.15). Für J_s aus (3.9) gilt gemäß (3.38) die Gleichheit $P_s J P_s^T = J_s$. Damit können wir für die asymptotische Verteilung einer zentrierten Version von $\hat{\mu}_s$ schreiben:

$$\sqrt{n} \left(\hat{\mu}_s - \mu(\theta_0, \gamma_0 + \frac{\delta}{\sqrt{n}}) \right) \xrightarrow{\mathcal{D}} \Lambda_s = c_s^T J_s^{-1} \begin{pmatrix} U \\ \pi_{S_s} V \end{pmatrix} + c_s^T L_s \delta,$$

wobei wir die Vektoren c_s und c sowie die Matrix L_s gemäß (3.11), (3.13) und (3.14) verwenden. Betrachten wir nun also $\sum_{s=1}^r w_s \Lambda_s$, so ergibt sich für den Erwartungswert

$$\nu(\xi, \delta, \theta_0, \gamma_0) = \mathbb{E} \left(\sum_{s=1}^r w_s \Lambda_s \right) = \sum_{s=1}^r w_s c_s^T L_s \delta.$$

Für die Varianz erhalten wir mit h_s aus (3.19):

$$\begin{aligned} \tau^2(\xi, \theta_0, \gamma_0) &= \mathbb{V}\text{ar} \left(\sum_{s=1}^r w_s \Lambda_s \right) = \sum_{s=1}^r \sum_{t=1}^r w_s w_t \text{Cov}(\Lambda_s, \Lambda_t) \\ &= \sum_{s=1}^r \sum_{t=1}^r w_s w_t \text{Cov} \left(c_s^T J_s^{-1} \begin{pmatrix} U \\ \pi_{S_s} V \end{pmatrix}, c_t^T J_t^{-1} \begin{pmatrix} U \\ \pi_{S_t} V \end{pmatrix} \right) \\ &= \sum_{s=1}^r \sum_{t=1}^r w_s w_t c_s^T J_s^{-1} P_s \text{Cov} \left(\begin{pmatrix} U \\ V \end{pmatrix}, \begin{pmatrix} U \\ V \end{pmatrix} \right) P_t^T J_t^{-1} c_t \\ &= \sum_{s=1}^r \sum_{t=1}^r w_s w_t h_s^T J h_t. \end{aligned}$$

□

Satz 3.7. Mit den Argumenten aus Pukelsheim (1993, Kapitel 11) und unter der Annahme, dass Integration und Differenziation vertauschbar sind, erfüllt gemäß Satz 2.5

beziehungsweise (2.7) ein Bayes-optimaler Versuchsplan ξ^* zur Model-Averaging Schätzung des Parameters μ die Ungleichheit

$$-D\Phi_{\text{mav}}(\xi^*)(\xi_x - \xi^*) = - \int_{\Theta \times \Gamma} D\Phi_{\text{mav}}(\xi^*, w, \delta, \theta, \gamma)(\xi_x - \xi^*) \pi(d\theta, d\gamma) \leq 0 \quad (3.41)$$

für alle $x \in \mathcal{X}$, wobei $D\Phi_{\text{mav}}(\xi^*, w, \delta, \theta, \gamma)(\xi_x - \xi^*)$ die Richtungsableitung (2.4) der Funktion Φ_{mav} aus (3.20) in ξ^* in Richtung $\xi_x - \xi^*$ bezeichnet. Weiterhin ist ξ_x das Dirac-Maß im Punkt $x \in \mathcal{X}$ aus (2.6).

Unter Verwendung der Kettenregel (Werner, 2018, Satz III.5.4) können wir nun die Richtungsableitung der Funktion Φ_{mav} bestimmen:

$$D\Phi_{\text{mav}}(\xi^*, w, \delta, \theta, \gamma)(\xi_x - \xi^*) = 2\nu(\xi^*)D_1(\xi^*, x, \delta, \theta, \gamma) + D_2(\xi^*, x, \theta, \gamma), \quad (3.42)$$

wobei $D_1(\xi^*, x, \delta, \theta, \gamma)$ die Richtungsableitung der Erwartungswertfunktion ν aus (3.17) und $D_2(\xi^*, x, \theta, \gamma)$ die Richtungsableitung der Varianzfunktion τ^2 aus (3.18) ist.

Zunächst betrachten wir die Richtungsableitung $D_1(\xi^*, x, \delta, \theta, \gamma)$, für welche wir

$$D_1(\xi^*, x, \delta, \theta, \gamma) = \sum_{s=1}^r w_s c^T DL_s(\xi^*, \theta, \gamma)(\xi_x - \xi^*) \delta \quad (3.43)$$

erhalten. Dabei bezeichnet $DL_s(\xi^*, \theta, \gamma)(\xi_x - \xi^*)$ die Richtungsableitung der Funktion L_s aus (3.14) und ist daher gegeben durch

$$DL_s(\xi^*, \theta, \gamma)(\xi_x - \xi^*) = P_s^T J_s^{-1}(\xi^*, \theta, \gamma_{S_s}) (P_s J(\xi_x, \theta, \gamma) - J_s(\xi_x, \theta, \gamma_{S_s}) J_s^{-1}(\xi^*, \theta, \gamma_{S_s}) P_s J(\xi^*, \theta, \gamma)) \begin{pmatrix} 0_{p \times q} \\ I_{q \times q} \end{pmatrix}. \quad (3.44)$$

Dabei finden sich die Matrizen P_s und $J_s(\xi_x, \theta, \gamma_{S_s})$ in (3.15) und (3.9) und $J(\xi_x, \theta, \gamma)$ ist die Informationsmatrix im größten Modell f_{wide} . In (3.44) haben wir genutzt, dass die Informationsmatrix $J_s(\xi)$ für ein beliebiges Kandidatenmodell $f_s \in \mathcal{S}$ linear ist in ξ , was bedeutet, dass

$$DJ_s(\xi^*, \theta, \gamma_{S_s})(\xi_x - \xi^*) = J_s(\xi_x, \theta, \gamma_{S_s}) - J_s(\xi^*, \theta, \gamma_{S_s}).$$

Weiterhin gilt

$$DJ_s^{-1}(\xi^*, \theta, \gamma_{S_s})(\xi_x - \xi^*) = J_s^{-1}(\xi^*, \theta, \gamma_{S_s}) - J_s^{-1}(\xi^*, \theta, \gamma_{S_s}) J_s(\xi_x, \theta, \gamma_{S_s}) J_s^{-1}(\xi^*, \theta, \gamma_{S_s}). \quad (3.45)$$

Die Kombination von (3.43) und (3.44) ergibt dann die Darstellung von $D_1(\xi^*, x, \delta, \theta, \gamma)$ in (3.23).

Die Richtungsableitung $D_2(\xi^*, x, \theta, \gamma)$ besitzt die Form

$$D_2(\xi^*, x, \theta, \gamma) = \sum_{s,t=1}^r w_s w_t (2Dh_s^T(\xi^*)(\xi_x - \xi^*)J(\xi^*, \theta, \gamma)h_t(\xi^*) + h_s^T(\xi^*)\{J(\xi^*, \theta, \gamma) + J(\xi_x, \theta, \gamma)\}h_t(\xi^*)), \quad (3.46)$$

wobei $Dh_s(\xi^*)(\xi_x - \xi^*)$ die Richtungsableitung von h_s in (3.19) für ein beliebiges Kandidatenmodell $f_s \in \mathcal{S}$ ist. Unter Beachtung von (3.45) ist die Richtungsableitung gegeben durch

$$Dh_s(\xi^*)(\xi_x - \xi^*) = h_s(\xi^*) - \tilde{h}_s(\xi^*, \xi_x), \quad (3.47)$$

mit

$$\tilde{h}_s(\xi^*, \xi) = P_s^T J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) J_s(\xi, \theta_0, \gamma_{0,S_s}) J_s^{-1}(\xi^*, \theta_0, \gamma_{0,S_s}) c_s,$$

siehe (3.25). Eine Verknüpfung von (3.46) und (3.47) liefert den Ausdruck für D_2 in (3.24). Die notwendige Bedingung in (3.22) resultiert durch Zusammenfügen von (3.41) und (3.42).

Es bleibt zu zeigen, dass in (3.22) Gleichheit für alle Trägerpunkte x des Versuchsplans ξ^* gilt. Angenommen, es existiert mindestens ein Trägerpunkt x_0 des Versuchsplans ξ^* , sodass in (3.22) echte Ungleichheit herrscht. Dann erhalten wir auch

$$\int_{\mathcal{X}} \int_{\Theta \times \Gamma} (-2\nu(\xi^*, \delta, \theta, \gamma) D_1(\xi^*, x, \delta, \theta, \gamma) - D_2(\xi^*, x, \theta, \gamma)) \pi(d\theta, d\gamma) \xi^*(dx) < 0.$$

Da aber $\int_{\mathcal{X}} J(\xi_x, \theta, \gamma) \xi^*(dx) = J(\xi^*, \theta, \gamma)$ und $\int_{\mathcal{X}} \tilde{h}_s(\xi^*, \xi_x) \xi^*(dx) = h_s(\xi^*)$ ist, resultiert andererseits

$$\int_{\mathcal{X}} D_1(\xi^*, x, \delta, \theta, \gamma) \xi^*(dx) = 0 \text{ und } \int_{\mathcal{X}} D_2(\xi^*, x, \theta, \gamma) \xi^*(dx) = 0,$$

sodass sich

$$\begin{aligned} & \int_{\mathcal{X}} \int_{\Theta \times \Gamma} -2\nu(\xi^*, \delta, \theta, \gamma) D_1(\xi^*, x, \delta, \theta, \gamma) - D_2(\xi^*, x, \theta, \gamma) \pi(d\theta, d\gamma) \xi^*(dx) \\ &= \int_{\Theta \times \Gamma} \left\{ -2\nu(\xi^*, \delta, \theta, \gamma) \int_{\mathcal{X}} D_1(\xi^*, x, \delta, \theta, \gamma) \xi^*(dx) - \int_{\mathcal{X}} D_2(\xi^*, x, \theta, \gamma) \xi^*(dx) \right\} \pi(d\theta, d\gamma) = 0 \end{aligned}$$

ergibt. Dies stellt einen Widerspruch dar und es folgt, dass in (3.22) Gleichheit für alle Trägerpunkte x des Versuchsplans ξ^* gelten muss. \square

3.6 Weitere Simulationsergebnisse

In diesem Abschnitt präsentieren wir weitere Simulationsergebnisse für die Beispiele aus Abschnitt 3.3.

3.6.1 Schätzung der $ED_{0.6}$ im sigmoidalen Emax-Modell

Dieser Abschnitt zeigt weitere Simulationsergebnisse für die Beispiele aus Abschnitt 3.3.1. In diesem haben wir Bayes-optimale Versuchspläne zur Model-Averaging Schätzung der $ED_{0.6}$ im sigmoidalen Emax-Modell bestimmt. In den Simulationen generieren wir Daten gemäß des Modells (3.29), wobei die $n = 150$ Beobachtungen jeweils gemäß der Versuchspläne ξ_A^* , ξ_B^* , ξ_1 und ξ_2 aus Abschnitt 3.3.1 erzeugt werden. Um zu demonstrieren, dass die Ergebnisse aus Abschnitt 3.3.1 repräsentativ sind, betrachten wir hier weitere Parameterkombinationen (ϑ, γ) . Die entsprechenden simulierten mittleren quadratischen Fehler sind in Tabelle 3.7 (gleiche Gewichtung: $w_s = 0.25, s = 1, \dots, 4$) und Tabelle 3.8 (unterschiedliche Gewichtung: $w_1 = 0.1, w_2 = 0.1, w_3 = 0.3$ and $w_4 = 0.5$) dargestellt. In der ersten und zweiten Spalte sind die Ergebnisse für die Model-Averaging Schätzung der $ED_{0.6}$ mit fester beziehungsweise glatter AIC-Gewichtung zu finden, während die dritte und vierte Spalte die Resultate für die Schätzer der $ED_{0.6}$ in dem durch das AIC gewählten Modell und in dem größten Modell präsentieren.

Wir beobachten in Tabelle 3.7, dass wir unter Verwendung der Model-Averaging Schätzer stets präzisere Schätzungen erhalten als mit dem Schätzer nach Modellselektion. In Zahlen beträgt die Reduktion des mittleren quadratischen Fehlers 14%-47% bei Gebrauch der festen (gleichen) Gewichtung beziehungsweise 7%-23% bei Nutzung der glatten AIC-Gewichtung im Vergleich zum Schätzer nach Modellselektion. Weiterhin verbessert die Verwendung der festen Gewichtung anstelle der datenbasierten glatten AIC-Gewichtung die Model-Averaging Schätzung. In allen betrachteten Szenarien besitzt der Model-Averaging Schätzer mit gleichen Gewichten kleinere mittlere quadratische Fehler als der Schätzer der $ED_{0.6}$ im größten Modell. Darüber hinaus verbessert der Model-Averaging Schätzer mit glatten AIC-Gewichten die Schätzung im größten Modell für beinahe alle untersuchten Parameterkombinationen und der Schätzer im mittels AIC gewählten Modells ist in mehr als der Hälfte der betrachteten Fälle präziser.

Auch in Tabelle 3.8 sind diese Phänomene zu sehen, jedoch ist der Unterschied zwischen dem Model-Averaging Schätzer basierend auf fester (unterschiedlicher) Gewichtung und glatter AIC-Gewichtung weniger deutlich. Zudem lassen einige Pa-

parameterkombinationen erkennen, dass ungleiche, feste Gewichte zu einem größeren mittleren quadratischen Fehler führen können als glatte AIC-Gewichte.

Abschließend vergleichen wir die optimalen Versuchspläne ξ_A^* und ξ_B^* mit den gleichverteilten Versuchsplänen ξ_1 und ξ_2 . Dieser Vergleich offenbart bei Gebrauch von ξ_A^* oder ξ_B^* eine Verringerung des mittleren quadratischen Fehlers des Model-Averaging Schätzers der $ED_{0.6}$ mit fester Gewichtung zwischen 21% und 56% gegenüber dem Versuchsplan ξ_1 und zwischen 28% und 54% gegenüber dem Versuchsplan ξ_2 . Für den Model-Averaging Schätzer basierend auf den glatten AIC-Gewichten können die optimalen Versuchspläne ξ_A^* und ξ_B^* den mittleren quadratischen Fehler um 8%-55% beziehungsweise 28%-53% verringern. Bei Schätzung der $ED_{0.6}$ im größten Modell beobachten wir eine Verringerung des mittleren quadratischen Fehlers um 11%-49% beziehungsweise 23%-54% bei Verwendung der optimalen Versuchspläne ξ_A^* beziehungsweise ξ_B^* . Bei der Schätzung in dem durch das AIC gewählten Modell reduzieren die optimalen Versuchspläne den mittleren quadratischen Fehler in den meisten betrachteten Fällen.

Parameter (ϑ, γ)	Versuchsplan	Schätzmethode			
		Feste Gew.	Glatte AIC-Gew.	Modellsel.	Gr. Modell
(1.81,0.79,0,1)	ξ_A^*	0.818	1.065	1.180	1.310
	ξ_1	1.339	1.526	1.660	1.631
	ξ_2	1.207	1.549	1.791	2.052
(1.81,0.79,0,1,1)	ξ_A^*	0.718	0.957	1.059	1.510
	ξ_1	1.238	1.413	1.535	1.852
	ξ_2	1.045	1.406	1.695	1.858
(1.81,0.79,0,2)	ξ_A^*	0.394	0.533	0.639	0.558
	ξ_1	0.788	0.823	0.915	1.097
	ξ_2	0.659	0.852	0.975	0.874
(1.81,0.79,0,1,2)	ξ_A^*	0.355	0.508	0.596	0.675
	ξ_1	0.810	0.913	1.017	1.267
	ξ_2	0.637	0.846	0.994	1.121
(1.81,1.79,0,2)	ξ_A^*	0.732	0.953	1.103	0.881
	ξ_1	1.374	1.570	1.767	1.591
	ξ_2	1.119	1.437	1.660	1.644
(1.81,1.79,0,1,2)	ξ_A^*	0.777	1.121	1.453	1.338
	ξ_1	1.166	1.384	1.532	1.500
	ξ_2	0.985	1.222	1.415	1.539
(1.81,1.79,0,3)	ξ_A^*	0.449	0.513	0.623	0.692
	ξ_1	0.988	1.144	1.250	1.116
	ξ_2	0.762	0.908	1.049	0.912
(1.81,1.79,0,1,3)	ξ_A^*	0.464	0.598	0.713	0.795
	ξ_1	0.932	1.182	1.314	1.202
	ξ_2	0.724	0.892	1.061	0.912

Tabelle 3.7: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.6}$ mit Gewichten $w_s = 0.25, s = 1, \dots, 4$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen. Innerhalb jeder Parameterkombination gehören die verschiedenen Zeilen zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_A^* zur Model-Averaging Schätzung der $ED_{0.6}$ aus (3.26). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Parameter (ϑ, γ)	Versuchsplan	Schätzmethode			
		Feste Gew.	Glatte AIC-Gew.	Modellsel.	Gr. Modell
(1.81,0.79,0,1)	ξ_B^*	0.864	0.849	1.012	1.500
	ξ_1	1.504	1.498	1.605	1.953
	ξ_2	1.382	1.450	1.631	2.020
(1.81,0.79,0.1,1)	ξ_B^*	0.914	0.937	1.112	1.441
	ξ_1	1.493	1.497	1.613	1.968
	ξ_2	1.306	1.310	1.491	1.967
(1.81,0.79,0,2)	ξ_B^*	0.540	0.536	0.600	0.535
	ξ_1	0.967	0.967	1.048	1.151
	ξ_2	0.834	0.861	1.004	1.067
(1.81,0.79,0.1,2)	ξ_B^*	0.476	0.502	0.582	0.562
	ξ_1	0.915	0.900	1.014	1.135
	ξ_2	0.869	0.949	1.067	1.103
(1.81,1.79,0,2)	ξ_B^*	0.904	0.873	1.038	1.173
	ξ_1	1.292	1.329	1.506	1.618
	ξ_2	1.362	1.338	1.611	1.522
(1.81,1.79,0.1,2)	ξ_B^*	0.875	0.931	1.091	1.070
	ξ_1	1.382	1.410	1.573	1.540
	ξ_2	1.350	1.368	1.599	1.588
(1.81,1.79,0,3)	ξ_B^*	0.516	0.532	0.619	0.780
	ξ_1	1.129	1.144	1.251	1.155
	ξ_2	0.836	0.813	0.927	1.086
(1.81,1.79,0.1,3)	ξ_B^*	0.578	0.560	0.615	0.606
	ξ_1	1.130	1.171	1.304	1.166
	ξ_2	0.800	0.851	1.023	1.049

Tabelle 3.8: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.6}$ mit Gewichten $w_1 = 0.1, w_2 = 0.1, w_3 = 0.3$ und $w_4 = 0.5$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen. Innerhalb jeder Parameterkombination gehören die verschiedenen Zeilen zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_B^* zur Model-Averaging Schätzung der $ED_{0.6}$ aus (3.30). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

3.6.2 Schätzung der AUC im logistischen Modell

In Ergänzung zu Abschnitt 3.3.2 präsentiert dieser Abschnitt weitere Simulationsergebnisse für die Model-Averaging Schätzung der Fläche unter der Kurve im logistischen Modell. Dabei erzeugen wir die Daten mit dem Modell (3.36) und generieren $n = 150$ Beobachtungen gemäß der Versuchspläne ξ_C^* , ξ_D^* , ξ_1 und ξ_2 . Um die Erkenntnisse aus Abschnitt 3.3.2 zu validieren, nutzen wir hier weitere Kombinationen der Parameter (ϑ, γ) und simulieren den mittleren quadratischen Fehler der Schätzer der Fläche unter der Kurve. Die Ergebnisse sind in Tabelle 3.9 (gleiche Modellgewichte: $w_s = 0.25, s = 1, \dots, 4$) und Tabelle 3.10 (unterschiedliche Modellgewichte: $w_1 = w_2 = w_3 = 0.1$ und $w_4 = 0.7$) zu finden. In deren erster Spalte präsentieren wir die Ergebnisse der Model-Averaging Schätzer mit festen Gewichten, während die zweite beziehungsweise dritte Spalte die jeweiligen Resultate des Model-Averaging Schätzers mit glatter AIC-Gewichtung beziehungsweise des Schätzers in dem Modell, das mittels AIC gewählt wurde, zeigt. Die vierte Spalte gibt die entsprechenden Ergebnisse für den Schätzer der $ED_{0.6}$ im größten Modell an.

Analog zu Abschnitt 3.3.2 ist festzustellen, dass der mittlere quadratische Fehler von den beiden Model-Averaging Verfahren stets kleiner ist als der des Schätzers im Anschluss an eine Modellsélection mittels des AIC (Verbesserung: 7%-26% mit gleicher Gewichtung, 1%-7% mit unterschiedlicher Gewichtung und 2%-10% mit glatter AIC-Gewichtung) und als der des Schätzers im größten Modell. Weiterhin liefert der Model-Averaging Schätzer mit gleichen Modellgewichten eine Verringerung des mittleren quadratischen Fehlers um 4%-18% (abhängig vom Versuchsplan und den verwendeten Parametern) verglichen mit dem Model-Averaging Schätzer mit glatter AIC-Gewichtung (siehe Tabelle 3.9). Dagegen führt die Verwendung der unterschiedlichen Gewichtung zu einem leichten Anstieg des mittleren quadratischen Fehlers (siehe Tabelle 3.10). Dies kann auf die in Abschnitt 3.3 diskutierte Problematik zurückzuführen sein, dass die feste, unterschiedliche Gewichtung nicht optimal sein muss für die Model-Averaging Schätzung in diesem Kontext. Es ist zu beachten, dass auch die gleiche Gewichtung nicht unbedingt optimal ist, jedoch scheint sie in diesem Kontext angemessener.

In Tabelle 3.9 beobachten wir, dass der Bayes-optimale Versuchsplan ξ_C^* den Model-Averaging Schätzer mit festen (gleichen) Gewichten in allen betrachteten Szenarien verbessert (Verringerung des mittleren quadratischen Fehlers um 2%-27%). Nutzen wir feste (ungleiche) Modellgewichte, so beträgt die Verbesserung durch Verwendung des optimalen Designs ξ_D^* nur 1%-16% und es gibt vereinzelte Parameterkombinationen, für die de facto keine Verbesserung auftritt (siehe Tabelle 3.10). Es ist zu

beachten, dass der Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der Fläche unter der Kurve so konstruiert ist, dass er für eine Menge an Parameterwerten (θ, γ) optimal ist. Dies kann zu einer geringeren Schätzgenauigkeit für einzelne Parameterwerte führen. Nutzen wir datenbasierte Gewichte (also glatte AIC-Gewichte oder Modellselektion), so erkennen wir in etwa der Hälfte der betrachteten Szenarien eine Verringerung des mittleren quadratischen Fehlers bei Nutzung von ξ_C^* anstelle von ξ_2 , während der Versuchsplan ξ_1 stets die größten mittleren quadratischen Fehler aufweist. Der Bayes-optimale Versuchsplan ξ_D^* , bestimmt für ungleiche, feste Modellgewichte, weist in fast allen untersuchten Fällen einen kleineren mittleren quadratischen Fehler als die Versuchspläne ξ_1 und ξ_2 auf, unabhängig von der verwendeten Schätzmethode. Bei Schätzung der $ED_{0.6}$ im größten Modell zeigt der gleichverteilte Versuchsplan ξ_2 über beide Tabellen hinweg in 75% der untersuchten Szenarien die kleinsten mittleren quadratischen Fehler.

Parameter (ϑ, γ)	Versuchsplan	Schätzmethode			
		Feste Gew.	Glatte AIC-Gew.	Modellsel.	Gr. Modell
(-1.73,4,0,1)	ξ_C^*	1.559	1.741	1.871	1.985
	ξ_1	1.886	1.963	2.030	2.168
	ξ_2	1.880	1.959	2.042	1.936
(-1.73,4,0.015,1)	ξ_C^*	1.503	1.658	1.802	2.069
	ξ_1	2.060	2.140	2.222	2.100
	ξ_2	1.831	1.917	1.981	1.840
(-1.73,4,0,0.833)	ξ_C^*	1.630	1.825	1.986	2.177
	ξ_1	2.042	2.139	2.230	2.092
	ξ_2	1.681	1.811	1.883	1.943
(-1.73,4,0.015,0.833)	ξ_C^*	1.659	1.880	2.074	2.071
	ξ_1	1.961	2.080	2.196	2.059
	ξ_2	1.687	1.763	1.838	1.847
(-1.73,5,0,0.833)	ξ_C^*	1.442	1.637	1.762	2.312
	ξ_1	1.671	1.815	1.925	2.036
	ξ_2	1.659	1.846	1.996	1.891
(-1.73,5,0.015,0.833)	ξ_C^*	1.517	1.773	1.953	2.380
	ξ_1	1.690	1.820	1.924	2.140
	ξ_2	1.629	1.764	1.884	1.950
(-1.73,5,0,0.667)	ξ_C^*	1.389	1.688	1.873	2.045
	ξ_1	1.672	1.823	1.955	2.269
	ξ_2	1.511	1.691	1.807	1.990
(-1.73,5,0.015,0.667)	ξ_C^*	1.421	1.687	1.839	2.243
	ξ_1	1.649	1.870	2.040	2.295
	ξ_2	1.626	1.792	1.907	1.943

Tabelle 3.9: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der AUC mit Gewichten $w_s = 0.25, i = 1, \dots, 4$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen. Innerhalb jeder Parameterkombination gehören die verschiedenen Zeilen zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_C^* zur Model-Averaging Schätzung der AUC aus (3.35). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Parameter (ϑ, γ)	Versuchsplan	Schätzmethode			
		Feste Gew.	Glatte AIC-Gew.	Modellsel.	Gr. Modell
(-1.73,4,0,1)	ξ_D^*	1.913	1.851	1.956	1.886
	ξ_1	2.159	2.128	2.213	2.269
	ξ_2	1.942	1.918	1.989	1.884
(-1.73,4,0.015,1)	ξ_D^*	1.890	1.843	1.951	2.039
	ξ_1	2.042	2.018	2.106	2.306
	ξ_2	1.935	1.912	1.959	2.090
(-1.73,4,0,0.833)	ξ_D^*	1.662	1.604	1.702	1.909
	ξ_1	1.964	1.934	2.025	2.084
	ξ_2	1.832	1.807	1.875	2.018
(-1.73,4,0.015,0.833)	ξ_D^*	1.764	1.723	1.835	1.978
	ξ_1	2.059	2.041	2.129	2.185
	ξ_2	1.841	1.801	1.883	1.886
(-1.73,5,0,0.833)	ξ_D^*	1.863	1.771	1.886	1.940
	ξ_1	1.881	1.818	1.930	2.194
	ξ_2	1.842	1.813	1.942	1.889
(-1.73,5,0.015,0.833)	ξ_D^*	1.689	1.617	1.761	1.970
	ξ_1	2.006	1.944	2.083	2.143
	ξ_2	1.700	1.670	1.815	1.994
(-1.73,5,0,0.667)	ξ_D^*	1.671	1.590	1.716	1.988
	ξ_1	1.833	1.769	1.925	2.086
	ξ_2	1.818	1.768	1.920	1.855
(-1.73,5,0.015,0.667)	ξ_D^*	1.745	1.665	1.816	2.079
	ξ_1	1.896	1.824	1.957	2.231
	ξ_2	1.649	1.626	1.779	1.900

Tabelle 3.10: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der AUC mit Gewichten $w_1 = w_2 = w_3 = 0.1$ und $w_4 = 0.7$ (erste Spalte), mit der glatten AIC-Gewichtung (2.24) (zweite Spalte), des Schätzers nach Modellselektion mit dem AIC (dritte Spalte) und des Schätzers im größten Modell (vierte Spalte). Die verschiedenen Zeilen entsprechen verschiedenen Parameterkombinationen. Innerhalb jeder Parameterkombination gehören die verschiedenen Zeilen zu verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan ξ_D^* zur Model-Averaging Schätzung der AUC aus (3.37). Zweite Zeile: Gleichverteilter Versuchsplan ξ_1 aus (3.27). Dritte Zeile: Gleichverteilter Versuchsplan ξ_2 aus (3.28).

Kapitel 4

Model-Averaging für nicht-verschachtelte Modelle

In Kapitel 3 haben wir mit dem Ziel optimale Versuchspläne zur Model-Averaging Schätzung eines Zielparameters μ zu bestimmen zunächst einen Spezialfall betrachtet, bei dem die Modelle verschachtelt sind und die Annahme lokaler Alternativen erfüllen. Hierdurch kann ein handlicher Ausdruck für die asymptotische Verteilung des Schätzers angegeben werden, welcher zur Bestimmung optimaler Versuchspläne genutzt werden kann. Diese Annahme kann sehr restriktiv sein, sodass beispielsweise Raftery und Zheng (2003) die Verwendung der Rahmenbedingungen aus Hjort und Claeskens (2003) anzweifeln. In der Praxis ist es zudem üblich, auch nicht-verschachtelte Modelle zu betrachten. Hierzu sei exemplarisch die *MCP-Mod* Methodik genannt, welche in Dosis-Wirkungs-Studien Anwendung findet und neben Modellwahlverfahren auch die Model-Averaging Schätzung über verschiedene (nicht-verschachtelte) Kandidatenmodelle nutzt (Bretz et al., 2005; Pinheiro et al., 2014). Aus diesem Grund entwickeln wir in diesem Kapitel ein Versuchsplanungskriterium für die Model-Averaging Schätzung unter festen Alternativen. Wir verwenden dabei das Wort „fest“ um das Konzept dieses Kapitels von dem Konzept der lokalen Alternativen aus Kapitel 3 abzugrenzen.

Aufbauend auf ersten Erkenntnissen aus Kapitel 3, welche zeigen, dass Model-Averaging Schätzer oftmals geringere mittlere quadratische Fehler aufweisen als Schätzer nach Modellselektion, zeigen wir in Abschnitt 4.1 zunächst, dass die Gültigkeit dieser Aussage stark von den betrachteten Kandidatenmodellen abhängt. In Abschnitt 4.2 leiten wir die asymptotische Verteilung von Model-Averaging Schätzern her, welche die Beobachtungen aus den Simulationsstudien erklärt. Basierend auf dieser Verteilung bestimmen wir ein Kriterium zur Bestimmung eines opti-

malen Versuchsplans, der den asymptotischen mittleren quadratischen Fehler des Model-Averaging Schätzers minimiert. Die Vorteile der Verwendung so konstruierter Versuchspläne illustrieren wir in Abschnitt 4.3 anhand von Simulationsstudien. Es zeigt sich, dass gerade bei Verwendung sehr unterschiedlicher Kandidatenmodelle die Nutzung datenabhängiger Gewichte zur Model-Averaging Schätzung vorteilhaft ist, weshalb wir in Abschnitt 4.4 eine adaptive Vorgehensweise vorschlagen, welche dieser Beobachtung Rechnung trägt. Abschließend schlagen wir in Abschnitt 4.5 eine Variante des in Abschnitt 4.2 entwickelten Versuchsplanungskriteriums vor, welche weniger Rechenkapazitäten benötigt und dadurch für die Anwendung attraktiv sein kann. Beweise zu den Aussagen dieses Kapitels finden sich in Abschnitt 4.6. Die in diesem Kapitel vorgestellten Ergebnisse finden sich zum Teil auch in Alhorn et al. (2019b).

4.1 Model-Averaging versus Modellselektion

Unter den Rahmenbedingungen aus Kapitel 2 betrachten wir Regressionsmodelle der Form (2.1). Dabei unterscheiden sich die Kandidatenmodelle jeweils in ihren Regressionsfunktionen η_s (mit Parameter ϑ_s) und wir gehen davon aus, dass insgesamt 4 verschiedene Regressionsfunktionen η_1, \dots, η_4 vorliegen. In diesem Abschnitt nehmen wir an, dass die Fehler normalverteilt sind, sodass sich für die Kandidatenmodelle die Dichten f_1, \dots, f_4 von Normalverteilungen mit Erwartungswertfunktion η_1, \dots, η_4 und Varianz σ^2 ergeben. Mittels der Kandidatenmodelle bestimmen wir jeweils Model-Averaging Schätzer gemäß (2.22), wobei wir zum einen gleiche Modellgewichte und zum anderen die glatten AIC-Gewichte aus (2.24) verwenden. Wir vergleichen diese mit dem Schätzer nach Modellselektion, welcher sich als Model-Averaging Schätzer mit den Gewichten aus (2.23) darstellen lässt.

Die folgende Simulationsstudie basiert auf einer Studie aus Pinheiro et al. (2006), welche die Dosis-Wirkungs-Beziehung von Beruhigungsmitteln untersucht. Dabei kann die Dosis des Medikamentes im Intervall $\mathcal{X} = [0, 150]$ variiert werden, wobei in der vorliegenden Studie die Patienten mit $k = 6$ verschiedenen Dosen $x_i \in \{0, 10, 25, 50, 100, 150\}$ behandelt wurden, auf welche sie möglichst gleich aufgeteilt wurden. Wir betrachten hier $n \in \{50, 100, 250\}$ Patienten und die Schätzung der $ED_{0.4}$, welche in (3.8) definiert ist.

Um die Unterschiede zwischen Modellselektion und Model-Averaging zu untersuchen, wählen wir im Folgenden zwei verschiedene Mengen von Kandidatenmodellen aus den in Tabelle 4.1 gegebenen Modellen aus.

Modell	Erwartungswertfunktion η_s	Parameterspezifikationen
Log-Linear (f_1)	$\eta_1(x_i, \vartheta_1) = \vartheta_{11} + \vartheta_{12} \log(x_i + \vartheta_{13})$	$\vartheta_1 = (0, 0.0797, 1)^T$
Emax (f_2)	$\eta_2(x_i, \vartheta_2) = \vartheta_{21} + \frac{\vartheta_{22}x_i}{\vartheta_{23}+x_i}$	$\vartheta_2 = (0, 0.467, 25)^T$
Exponentiell (f_3)	$\eta_3(x_i, \vartheta_3) = \vartheta_{31} + \vartheta_{32} \exp(x_i/\vartheta_{33})$	$\vartheta_3 = (-0.08265, 0.08265, 85)^T$
Quadratisch (f_4)	$\eta_4(x_i, \vartheta_4) = \vartheta_{41} + \vartheta_{42}x_i + \vartheta_{43}x_i^2$	$\vartheta_4 = (0, 0.00533, -0.00002)^T$

Tabelle 4.1: Verwendete Modelle und Parameter zum Vergleich von Model-Averaging und Modellselektion bei nicht-verschachtelten Modellen.

Die erste Menge

$$\mathcal{S}_1 = \{f_1, f_2, f_4\} \quad (4.1)$$

enthält das log-lineare, das Emax und das quadratische Modell, während die zweite Menge

$$\mathcal{S}_2 = \{f_1, f_2, f_3\} \quad (4.2)$$

das log-lineare, das Emax und das exponentielle Modell enthält. Dabei dient die erste Menge als Prototyp für eine Menge ähnlicher Kandidatenmodelle, während die Modelle in \mathcal{S}_2 eine eher unterschiedliche Form aufweisen. Diese Unterschiede sind in Abbildung 4.1 illustriert. In der linken Grafik ist das quadratische Modell mit den Parametern aus Tabelle 4.1 dargestellt, sowie die besten Approximationen der Erwartungswertfunktion durch ein log-lineares (f_1) und ein Emax Modell (f_2) bezüglich des Kullback-Leibler Abstandes (2.13). Es wird deutlich, dass alle Funktionen eine ähnliche Form haben. Dies drückt sich auch in den einzelnen $ED_{0.4}$ aus, welche 32.581, 32.261 und 33.810 für das log-lineare, das Emax und das quadratische Modelle beträgt. Analog zeigt die rechte Grafik in Abbildung 4.1 das exponentielle Modell mit den Parametern aus Tabelle 4.1 und seine besten Approximationen durch das log-lineare und das Emax Modell. Hier zeigen sich nun größere Unterschiede zwischen den Modellen, welche sich auch in den unterschiedlichen $ED_{0.4}$ von 58.116, 42.857 und 91.547 für das log-lineare, das Emax und das exponentielle Modell widerspiegeln.

Die Ergebnisse, die wir in diesem Kapitel präsentieren, basieren jeweils auf 1000 Simulationsdurchläufen, in denen jeweils n Beobachtungen der Form

$$y_{ij}^{(l)} = \eta_s(x_i, \vartheta_s) + \varepsilon_{ij}^{(l)}, i = 1, \dots, k, j = 1, \dots, n_i, \quad (4.3)$$

erzeugt werden. Dabei sind $\varepsilon_{ij}^{(l)}$ unabhängige, zentrierte normalverteilte Zufallsvariablen mit Varianz $\sigma^2 = 0.1$ und η_s ist eine der Funktionen η_1, \dots, η_4 (jeweils mit

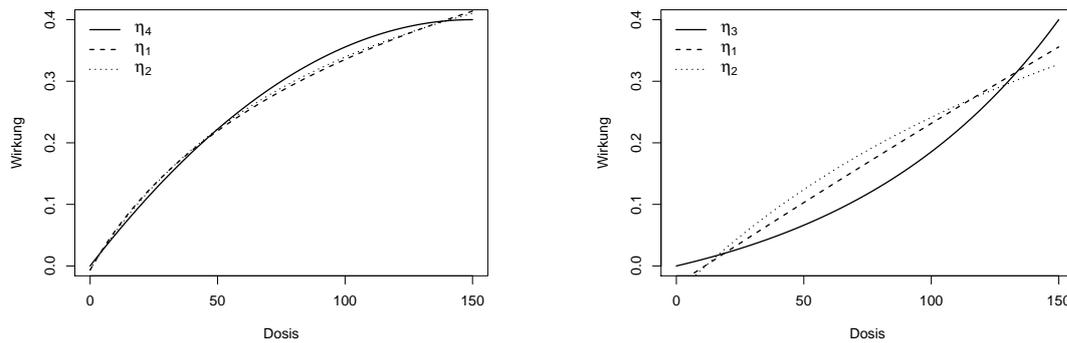


Abbildung 4.1: Links: Quadratisches Modell (durchgezogene Linie) und die besten Approximationen durch das log-lineare (gestrichelte Linie) und das Emax Modell (gepunktete Linie) bezüglich des Kullback-Leibler-Abstandes. Rechts: Exponentielles Modell (durchgezogene Linie) und die besten Approximationen durch das log-lineare (gestrichelte Linie) und das Emax Modell (gepunktete Linie) bezüglich des Kullback-Leibler-Abstandes.

der Parameterspezifikation aus Tabelle 4.1). In jedem Simulationsdurchlauf schätzen wir den Parameter $\mu = ED_{0.4}$ mittels Model-Averaging mit gleicher Gewichtung, mit glatter AIC-Gewichtung (2.24) und mittels des Schätzers nach Modellselektion mit dem AIC. Mit den resultierenden Schätzern bestimmen wir jeweils den empirischen mittleren quadratischen Fehler. Wir berechnen also mittels der Schätzungen $\hat{\mu}_{\text{mav}}^{(l)}$ aus (2.22) aus jedem Simulationsdurchlauf den Fehler über

$$\text{MSE}(\xi) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\mu}_{\text{mav}}^{(l)} - \mu_{\text{true}})^2, \quad (4.4)$$

wobei μ_{true} jeweils die $ED_{0.4}$ in dem in (4.3) verwendeten Modell ist.

Die mit den unterschiedlichen Mengen von Kandidatenmodellen simulierten mittleren quadratischen Fehler sind jeweils in Tabelle 4.2 und Tabelle 4.3 dargestellt. Diese zeigen die Ergebnisse unter Verwendung des Model-Averaging Schätzers mit gleichen Gewichten (linke Spalte), unter Verwendung von glatten AIC-Gewichten (mittlere Spalte) und unter Verwendung des Schätzers nach Modellselektion (rechte Spalte). Dabei nutzen wir in (4.3) unterschiedliche Modelle zur Generierung der Daten. Die zugehörigen Simulationsergebnisse sind in den unterschiedlichen Zeilen der entsprechenden Tabellen dargestellt. Zudem geben wir die Ergebnisse unter Verwendung unterschiedlicher Stichprobengrößen n an. Die fettgedruckten Zahlen weisen jeweils auf die Schätzmethode mit dem kleinsten mittleren quadratischen Fehler innerhalb jeder Zeile hin.

4.1.1 Modelle von ähnlicher Form

In diesem Abschnitt setzen wir uns zunächst mit den Simulationsergebnissen für die Menge \mathcal{S}_1 von Kandidatenmodellen mit ähnlicher Form auseinander (siehe Tabelle 4.2). Ist das datengenerierende Modell auch Teil der Menge \mathcal{S}_1 (verwenden wir also f_1 , f_2 oder f_4), so zeigt Tabelle 4.2, dass die Model-Averaging Schätzung mit gleicher Gewichtung sehr präzise ist. Bis auf zwei Fälle ist der mittlere quadratische Fehler dieser Schätzmethode um 10%-25% geringer als der des Model-Averaging Schätzers mit glatter AIC-Gewichtung. Zudem ist der mittlere quadratische Fehler des Schätzers nach Modellselektion in nahezu allen betrachteten Fällen deutlich größer als der mittlere quadratische Fehler der beiden Model-Averaging Schätzer, insbesondere bei kleiner Stichprobengröße. Dies ist durch die zusätzliche Variabilität, die mit datenbasierten Gewichten einhergeht zu begründen. Ist beispielsweise bei einer Stichprobengröße von $n = 50$ das log-lineare Modell f_1 das datengenerierende Modell, so ist der mittlere quadratische Fehler des Model-Averaging Schätzers mit festen Gewichten 437.045. Bei Verwendung von glatten AIC-Gewichten beträgt dieser dagegen 498.323 und bei Schätzung in dem mittels AIC gewählten Modell sogar 758.978. Dabei teilen sich diese jeweils in eine Varianz von 235.152, 337.609 beziehungsweise 599.678 und eine quadrierte Verzerrung von 201.894, 160.714 und 159.300 auf. Dementsprechend kann die Verzerrung des Schätzers durch Verwendung von datenbasierten Gewichten verringert werden, da hierdurch ein größeres Gewicht auf das „richtige“ Modell gelegt wird. Mit steigender Stichprobengröße sinkt die Varianz der Schätzer, sodass Schätzer basierend auf zufälligen Gewichten für größere Stichproben präziser werden und teils kleinere mittlere quadratische Fehler besitzen als Model-Averaging Schätzer mit festen Gewichten. Insgesamt ist jedoch festzuhalten, dass Model-Averaging präziser ist als Modellselektion, sofern das „wahre“ Modell unter den Kandidatenmodellen ist und sich die Menge ebendieser ähnlich ist. Insbesondere Model-Averaging Schätzer mit gleichen Gewichten liefern sehr gute Ergebnisse. Dies stimmt mit den Beobachtungen aus Kapitel 3 überein und ähnliche Ergebnisse finden sich in Schorning et al. (2016) und Buatois et al. (2018) im Kontext von Dosis-Wirkungs-Studien. Zudem beobachten Chen et al. (2018) ähnliche Effekte für AIC-basierte Gewichte im Zusammenhang mit multinomialen Logit Modellen.

Ist dagegen keines der Modelle aus der Menge der Kandidatenmodelle das „wahre“ Modell, so ändert sich dieses Verhalten. Dies ist im unteren Teil von Tabelle 4.2 dargestellt, welcher die mittleren quadratischen Fehler unter Nutzung des exponentiellen Modells f_3 zur Datenerzeugung enthält, welches nicht in \mathcal{S}_1 enthalten ist.

Modell	Stichprobengr.	Schätzmethode		
		Gleiche Gew.	Glatte AIC-Gew.	Modellselektion
f_1	$n = 50$	437.045	498.323	758.978
	$n = 100$	223.291	218.99	285.062
	$n = 250$	111.973	82.713	78.371
f_2	$n = 50$	286.638	329.904	515.32
	$n = 100$	189.785	203.796	251.836
	$n = 250$	62.792	64.854	66.54
f_4	$n = 50$	276.037	361.101	669.873
	$n = 100$	190.662	244.558	391.443
	$n = 250$	92.653	109.852	139.859
f_3	$n = 50$	1503.903	1372.31	1381.033
	$n = 100$	1109.622	856.484	729.912
	$n = 250$	864.163	398.144	255.604

Tabelle 4.2: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0,4}$ mit gleichen Gewichten (linke Spalte), mit glatten AIC-Gewichten (mittlere Spalte) und des Schätzers nach Modellselektion mit dem AIC (rechte Spalte). Die Menge der Kandidatenmodelle ist $\mathcal{S}_1 = \{f_1, f_2, f_4\}$. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen aus Tabelle 4.1 und verschiedenen Stichprobengrößen n .

Hier ist zu sehen, dass die Model-Averaging Schätzer mit glatten AIC-Gewichten stets besser sind als die Model-Averaging Schätzer mit gleicher Gewichtung. Zudem sind die Schätzer nach Modellselektion mit dem AIC bei steigendem Stichprobenumfang präziser als die Model-Averaging Schätzer. Dieses Verhalten ist durch die unterschiedlichen Formen der Regressionsfunktionen zu erklären, welche in Abbildung 4.1 dargestellt sind. Durch eine entsprechende Wahl der Parameter kann das quadratische Modell die Form des exponentiellen Modells annehmen, während selbst die besten Approximationen durch das log-lineare und das Emax Modell weiterhin eine andere Gestalt aufweisen als das exponentielle Modell (siehe rechte Grafik in Abbildung 4.1). Ein Einschluss dieser Modelle in die Model-Averaging Schätzung führt dementsprechend zu einer großen Verzerrung, die deutlich reduziert werden kann durch glatte AIC-Gewichte oder durch Modellselektion. Ist beispielsweise $n = 100$, so beträgt die quadrierte Verzerrung des Model-Averaging Schätzers mit gleichen Modellgewichten 981.631, während sie 328.634 beziehungsweise 69.465 für den Model-Averaging Schätzer mit glatten AIC-Gewichten beziehungsweise den Schätzer nach Modellselektion beträgt.

4.1.2 Modelle von unterschiedlicher Form

Wir betrachten nun die Menge \mathcal{S}_2 aus (4.2), welche als Beispiel für eine Menge von Kandidatenmodellen von unterschiedlicher Form dient und das log-lineare, das Emax und das exponentielle Modell enthält. Die simulierten mittleren quadratischen Fehler der drei Schätzungen der $ED_{0.4}$ sind in Tabelle 4.3 zu finden. Der obere Teil der Tabelle enthält die Szenarien, in denen das „wahre“ Modell unter den Kandidatenmodellen in \mathcal{S}_2 ist. Im Gegensatz zu Abschnitt 4.1.1 zeigt sich hier lediglich ein Szenario, in dem der Model-Averaging Schätzer mit gleichen Gewichten den kleinsten mittleren quadratischen Fehler innerhalb der drei Schätzmethoden besitzt. Je größer der Stichprobenumfang ist, desto deutlicher tritt die Verbesserung, die sich durch Verwendung von datenbasierten anstelle von festen Gewichten ergibt, zutage. Eine Erklärung für dieses Verhalten ist, dass die Verwendung von gleichen Modellgewichten bei einer Menge von Kandidatenmodellen mit sehr unterschiedlicher Form eine große Verzerrung mit sich bringt. Model-Averaging Schätzer mit glatten AIC-Gewichten und Schätzer nach Modellselektion hingegen legen mehr Gewicht auf das „wahre“ Modell, insbesondere wenn die Stichprobe groß ist. Auch wenn der Schätzer nach Modellselektion eine größere Varianz besitzt als der Model-Averaging Schätzer mit glatten AIC-Gewichten, ist dieser für große Stichproben effizienter. Dies ist darin begründet, dass die Varianz der Schätzung mit dem Stichprobenumfang sinkt, wodurch die Verzerrung des Schätzers den mittleren quadratischen Fehler ab einer gewissen Stichprobengröße dominiert.

Ist das „wahre“ Modell das quadratische Modell $f_4 \notin \mathcal{S}_2$, so besitzt der Model-Averaging Schätzer mit gleichen Gewichten für kleine und mittlere Stichprobengrößen den kleinsten mittleren quadratischen Fehler. In diesen Szenarien weist der Schätzer nach Modellselektion einen wesentlich größeren mittleren quadratischen Fehler auf als die Model-Averaging Verfahren. Die Unterschiede verringern sich jedoch mit wachsender Stichprobengröße. Insbesondere weisen für $n = 250$ der Model-Averaging Schätzer mit glatten AIC-Gewichten und der Schätzer nach Modellselektion kleinere mittlere quadratische Fehler auf als der Model-Averaging Schätzer mit gleichen Gewichten.

Insgesamt lässt sich anhand der Ergebnisse aus Abschnitt 4.1.1 und Abschnitt 4.1.2 bilanzieren, dass anders als in Kapitel 3 und weiteren Simulationsstudien in der Literatur, Model-Averaging Schätzer nicht zwangsläufig effizienter sind als Schätzer nach Modellselektion. Insbesondere sind Model-Averaging Schätzer mit gleichen Modellgewichten nur dann präziser als die beiden verwendeten datenbasierten Verfahren, wenn die Kandidatenmodelle „ähnlich“ sind. Besitzen die Kandidatenmodelle da-

Modell	Stichprobengr.	Schätzmethode		
		Gleiche Gew.	Glatte AIC-Gew.	Modellselektion
f_1	$n = 50$	834.295	553.427	776.311
	$n = 100$	712.404	340.254	353.707
	$n = 250$	524.518	48.587	38.591
f_2	$n = 50$	640.706	505.054	669.285
	$n = 100$	517.963	267.967	286.272
	$n = 250$	394.536	65.805	53.424
f_3	$n = 50$	1076.154	1141.476	1427.441
	$n = 100$	871.362	766.140	802.763
	$n = 250$	802.196	480.641	399.839
f_4	$n = 50$	288.091	486.501	852.377
	$n = 100$	208.628	298.315	419.651
	$n = 250$	162.689	138.331	142.673

Tabelle 4.3: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0,4}$ mit gleichen Gewichten (linke Spalte), mit glatten AIC-Gewichten (mittlere Spalte) und des Schätzers nach Modellselektion mit dem AIC (rechte Spalte). Die Menge der Kandidatenmodelle ist $\mathcal{S}_2 = \{f_1, f_2, f_3\}$. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen aus Tabelle 4.1 und verschiedenen Stichprobengrößen n .

gegen sehr unterschiedliche Formen, so sind Model-Averaging Schätzer mit glatten AIC-Gewichten oder der Schätzer nach Modellselektion der Variante mit festen Gewichten vorzuziehen. Diese Beobachtungen lassen sich anhand der theoretischen Resultate im folgenden Kapitel erklären. Wir zeigen zudem anhand von Simulationen, dass die auf Basis dieser Theorie entwickelten Versuchspläne die Schätzgenauigkeit aller drei Schätzmethode stark verbessern können.

4.2 Asymptotische Verteilung und optimale Versuchsplanung

In diesem Abschnitt leiten wir asymptotische Eigenschaften von Model-Averaging Schätzern mit festen Modellgewichten für die allgemeinen Annahmen (A1)-(A6) aus Kapitel 2 her. Insbesondere verzichten wir also auf die Annahme verschachtelter Modelle und treffen nur milde Annahmen an das wahre Modell g . Dieser Rahmen erlaubt die Beobachtungen aus Abschnitt 4.1 zu erklären, insbesondere die Domi-

nanz der Verzerrung im mittleren quadratischen Fehler für große Stichprobengrößen n . Diese Ergebnisse nutzen wir im Weiteren, um optimale Versuchspläne zu bestimmen, die den asymptotischen mittleren quadratischen Fehler des Model-Averaging Schätzers minimieren.

4.2.1 Asymptotische Verteilung

Hjort und Claeskens (2003) bestimmen die asymptotische Verteilung von frequentistischen Model-Averaging Schätzern unter Verwendung der Annahme lokaler Alternativen. Dieses in Kapitel 3 verwendete Konzept setzt voraus, dass die wahre Dichte g in einem größten parametrischen Kandidatenmodell enthalten ist. Sämtliche Kandidatenmodelle sind Teilmengen von diesem und die Abweichungen der Parameter sind so beschränkt, dass sie maximal von der Größe $n^{-1/2}$ sind (siehe (3.2)). Hierdurch kann in Satz 3.4 ein handlicher Ausdruck für den mittleren quadratischen Fehler der Grenzverteilung des Model-Averaging Schätzers angegeben werden, da sowohl Verzerrung als auch Varianz von der Ordnung $\mathcal{O}(1/n)$ sind. Dieses Kapitel verzichtet auf derartige Annahmen, sodass lediglich die Annahmen aus Kapitel 2 Anwendung finden.

Im Folgenden bezeichne die Zufallsvariable Y den Ausgang eines Experimentes unter Versuchsbedingung x . Wir gehen analog zu Kapitel 2 davon aus, dass r verschiedene Kandidatenmodelle mit Dichten

$$f_1(y | x, \theta_1), \dots, f_r(y | x, \theta_r) \quad (4.5)$$

vorliegen. Allgemein muss die Dichte g des wahren Modells nicht in der Menge der Kandidatenmodelle (4.5) enthalten sein. Die in Abschnitt 4.1 betrachteten Regressionsmodelle der Form (2.1) sind ein Spezialfall der hier verwendeten Theorie. Wie bereits in Kapitel 2 erläutert, existiert in jedem der Kandidatenmodelle ein (Quasi-)Maximum-Likelihood Schätzer $\hat{\theta}_{n,s}$, welcher stark konsistent für den Wert $\theta_{s,g}^*(\xi) = \theta_s^*(\xi)$ aus (2.14) ist. Weiterhin ist eine zentrierte Version dieses Schätzers gemäß Satz 2.11 asymptotisch normalverteilt. Mit diesen Ergebnissen können wir nun die asymptotische Verteilung von Schätzern der Form (2.22) angeben.

Satz 4.1. *Unter den Annahmen (A1)-(A6) aus Kapitel 2 gilt für den Model-Averaging Schätzer aus (2.22) mit festen Modellgewichten*

$$\sqrt{n} \left(\hat{\mu}_{mav} - \sum_{s=1}^r w_s \mu_s(\theta_s^*(\xi)) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma_w^2(\theta^*(\xi)) \right),$$

wobei

$$\sigma_w^2(\theta^*(\xi)) = \sum_{s=1}^r \sum_{t=1}^r w_s w_t \left(\frac{\partial \mu_s(\theta_s^*(\xi))}{\partial \theta_s} \right)^T A_s^{-1}(\theta_s^*(\xi)) B_{st}(\theta_s^*(\xi), \theta_t^*(\xi)) A_t^{-1}(\theta_t^*(\xi)) \frac{\partial \mu_t(\theta_t^*(\xi))}{\partial \theta_t}. \quad (4.6)$$

Diese Grenzverteilung ist wesentlich komplexer als die Verteilung in (3.16) unter der Annahme lokaler Alternativen. Vor allem die Abhängigkeit der Parameter $\theta_s^*(\xi)$ vom Versuchsplan führt dazu, dass die Berechnung der Varianz in (4.6) sehr aufwändig ist. Satz 4.1 zeigt, dass der Model-Averaging Schätzer nicht unverzerrt ist, sofern nicht $\sum_{s=1}^r w_s \mu_s(\theta_s^*(\xi)) = \mu_{\text{true}}$ gilt. Aus diesem Grund erläutern wir im Folgenden ein Versuchsplanungskriterium, das den asymptotischen mittleren quadratischen Fehler des Model-Averaging Schätzers minimiert. Im Gegensatz zu Kapitel 3 hängt hier die Verzerrung des Schätzers nicht von der Stichprobengröße ab, während die Varianz von der Ordnung $\mathcal{O}(1/n)$ ist.

4.2.2 Optimale Versuchsplanung

In Kapitel 3 haben wir Versuchspläne bestimmt, die den mittleren quadratischen Fehler des Model-Averaging Schätzers bei verschachtelten Kandidatenmodellen unter der Annahme lokaler Alternativen minimieren und demonstriert, dass diese die Schätzgenauigkeit im Vergleich zu gleichverteilten Versuchsplänen, die in Dosis-Wirkungs-Studien gebräuchlich sind, stark verbessern können. Mit dem Resultat aus Satz 4.1 können wir dieses Versuchsplanungskriterium auf Situationen erweitern, in denen die Kandidatenmodelle nicht verschachtelt sind und selbst das wahre Modell nicht unter den Kandidatenmodellen sein muss. Wir betrachten im Folgenden also das Versuchsplanungskriterium

$$\Phi_{\text{mav}}(\xi, g) = \frac{1}{n} \sigma_w^2(\theta^*(\xi)) + \left(\sum_{s=1}^r w_s \mu_s(\theta_s^*(\xi)) - \mu_{\text{true}} \right)^2 \approx \text{MSE}(\hat{\mu}_{\text{mav}}), \quad (4.7)$$

wobei μ_{true} der Zielparameter im wahren Modell mit Dichte g ist und $\sigma_w^2(\theta^*(\xi))$ und $\theta_s^*(\xi)$ in (4.6) beziehungsweise (2.14) definiert sind. Dieses Kriterium hängt dementsprechend von der wahren Verteilung über μ_{true} sowie die Parameter $\theta_s^*(\xi) = \theta_{s,g}^*(\xi)$ ab, mittels derer die beste Approximation des wahren Modells durch ein Kandidatenmodell bezüglich des Kullback-Leibler-Abstands erreicht wird.

Um nun einen Zielparameter μ möglichst präzise mittels Model-Averaging zu schätzen, sollte ein „guter“ Versuchsplan möglichst kleine Werte der Funktion $\Phi_{\text{mav}}(\xi, g)$ aus (4.7) erzielen.

Definition 4.2. Für eine endliche Menge \mathcal{S} von Kandidatenmodellen f_s , $s = 1, \dots, r$, welche die obigen Annahmen erfüllt, und feste Gewichte w_s ($s = 1, \dots, r$) heißt ein Versuchsplan ξ^* *lokal optimaler Versuchsplan zur Model-Averaging Schätzung des Parameters μ* , falls er die Funktion $\Phi_{\text{mav}}(\xi, g)$ in (4.7) in der Klasse aller approximativen Versuchspläne auf \mathcal{X} minimiert, das heißt falls

$$\Phi_{\text{mav}}(\xi^*, g) = \min_{\xi \in \Xi} \Phi_{\text{mav}}(\xi, g).$$

Auch hier verwenden wir den Begriff lokal analog zu Chernoff (1953), da das Optimalitätskriterium in Definition 4.2 von der unbekanntem Dichte $g(y | x)$ abhängt. Wie in Kapitel 3 verwenden wir einen bayesianischen Ansatz, um ein von der unbekanntem Dichte unabhängiges Versuchsplanungskriterium zu erhalten. Bezeichne \mathcal{G} eine endliche Menge möglicher Dichten und sei π ein Wahrscheinlichkeitsmaß auf \mathcal{G} .

Definition 4.3. Unter den Annahmen aus Definition 4.2 heißt ein Versuchsplan ξ^* *Bayes-optimal zur Model-Averaging Schätzung des Parameters μ hinsichtlich der A-priori Verteilung π* , falls er die Funktion

$$\Phi_{\text{mav}}^\pi(\xi) = \int_{\mathcal{G}} \Phi_{\text{mav}}(\xi, g) d\pi(g) \quad (4.8)$$

minimiert, das heißt, falls

$$\Phi_{\text{mav}}^\pi(\xi^*) = \min_{\xi \in \Xi} \Phi_{\text{mav}}^\pi(\xi).$$

Wir gehen dabei stets davon aus, dass das Integral in (4.8) existiert.

Allgemein können wir die Menge \mathcal{G} unabhängig von der Menge der Kandidatenmodelle konstruieren. Liegt jedoch kein spezifisches Vorwissen über das wahre Modell vor, so kann eine Menge \mathcal{G} auf Grundlage der Kandidatenmodelle gebildet werden. Dazu bezeichnen wir die Menge der Dichten in (4.5) mit \mathcal{S} . Jedes der Kandidatenmodelle hängt von einem unbekanntem Parameter θ_s ab und wir nutzen die Bezeichnung $\mathcal{F}_{f_s} \subset \Theta_s$ für eine Menge möglicher wahrer Parameterwerte für das Modell f_s . Sei nun π_2 eine A-priori Verteilung auf der Menge \mathcal{S} und für jedes $f_s \in \mathcal{S}$ durch $\pi_1(\cdot | f_s)$ eine A-priori Verteilung auf \mathcal{F}_{f_s} gegeben. Dann definieren wir die Menge der wahren Modelle durch $\mathcal{G} = \{(g, \theta) : g \in \mathcal{S}, \theta \in \mathcal{F}_g\}$ und eine A-priori Verteilung auf dieser durch

$$d\pi(g, \theta) = d\pi_1(\theta | g) d\pi_2(g). \quad (4.9)$$

Dadurch können wir das Kriterium (4.8) umschreiben zu

$$\Phi_{\text{mav}}^\pi(\xi) = \int_{\mathcal{S}} \int_{\mathcal{F}_g} \Phi_{\text{mav}}(\xi, g) d\pi_1(\theta | g) d\pi_2(g). \quad (4.10)$$

In der Simulationsstudie im folgenden Abschnitt sind sowohl die Menge \mathcal{S} als auch die Menge \mathcal{F}_{f_s} (für jedes $f_s \in \mathcal{S}$) endlich, sodass auch die Menge \mathcal{G} endlich ist.

Bemerkung. Die Bezeichnung der optimalen Versuchspläne in diesem Kapitel ist, um die Notation kurz zu halten, identisch zu den Bezeichnungen aus Kapitel 3. Sofern nicht anders angegeben, verwenden wir in diesem Kapitel die Bezeichnungen stets gemäß Definition 4.2 und Definition 4.3.

Die Bestimmung eines optimalen Versuchsplans ist aufgrund des komplexen Ausdrucks in Definition 4.3 lediglich numerisch möglich. Durch die Abhängigkeit der Parameterwerte $\theta_s^*(\xi)$ aus (2.14) vom optimalen Versuchsplan ist die Herleitung einer notwendigen Bedingung mit Hilfe von Satz 2.5 deutlich aufwändiger als in Kapitel 3, sodass wir diesen Ansatz in dieser Arbeit nicht weiter verfolgen. Daher liegt kein Instrument vor, um die Optimalität eines numerisch bestimmten Versuchsplans zu überprüfen. Für die in Abschnitt 4.5 hergeleitete Variante des hier angegebenen Versuchsplanungskriteriums jedoch geben wir eine notwendige Bedingung für das Vorliegen eines Minimums an.

Das in diesem Abschnitt vorgeschlagene Versuchsplanungskriterium basiert - ebenso wie jenes aus Kapitel 3 - auf der Annahme, dass die Modellgewichte fest gewählt sind. Hjort und Claeskens (2003) können unter der Annahme lokaler Alternativen einen Ausdruck für die asymptotische Verteilung von Model-Averaging Schätzern mit zufälligen Gewichten angeben. Diese ist eine Mischung zweier Normalverteilungen und sie schlagen basierend hierauf Möglichkeiten vor, mittels numerischer Integration oder anhand von Simulationen die Verteilung von Model-Averaging Schätzern mit zufälligen Gewichten numerisch zu bestimmen. Grundlegend für ihren Ansatz ist dabei, dass sich die Grenzverteilung der Modellgewichte sowie die asymptotische Verteilung des Schätzers des Zielparameters als Funktion der gleichen Zufallsvariablen schreiben lassen. Die Angabe der asymptotischen Verteilung eines Model-Averaging Schätzers mit zufälligen Gewichten ist unter den allgemeinen Annahmen in diesem Kapitel nicht für beliebige Gewichte $w_s(Y_n)$ möglich. Unter der Bedingung jedoch, dass die zufälligen Gewichte $w_s(Y_n)$ gegen Konstanten w_s konvergieren, kann mit dem Lemma von Slutsky eine solche Verteilung angegeben werden. Dies illustrieren wir im folgenden Beispiel für die glatten AIC-Gewichte aus (2.24).

Wir zeigen in Abschnitt 4.3 anhand von Simulationen, dass die Verwendung von Versuchsplänen, die gemäß des hier vorgeschlagenen Kriteriums basierend auf festen Gewichten konstruiert sind, zu einer Verbesserung der Schätzgenauigkeit von Model-Averaging Schätzern mit glatten AIC-Gewichten sowie von Schätzern nach

Modellselektion führt. Zudem schlagen wir in Abschnitt 4.4 eine adaptive Vorgehensweise vor, welche datenbasierte Gewichte berücksichtigt.

Beispiel 4.4. Wir betrachten Kandidatenmodelle mit Dichten f_1, \dots, f_r und die glatten AIC-Gewichte aus (2.24). Dann gilt nach White (1981, Lemma 3.1) für das Maximum der Quasi-log-Likelihood im Kandidatenmodell f_s :

$$\ell_{n,s}(\hat{\theta}_{n,s} | Y_n) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \log f_s(Y_{ij} | x_i, \hat{\theta}_{n,s}) \xrightarrow{a.s.} \sum_{i=1}^k \xi_i \mathbb{E}_{g(\cdot | x_i)} [\log f_s(Y_{ij} | x_i, \theta_{s,g}^*(\xi))].$$

Damit folgt für das AIC aus (2.20) mit dem Continuous Mapping Theorem

$$\text{AIC}(f_s | Y_n) \xrightarrow{a.s.} 2 \sum_{i=1}^k \xi_i \mathbb{E}_{g(\cdot | x_i)} [\log f_s(Y_{ij} | x_i, \theta_{s,g}^*(\xi))] - 2p_s =: a_s$$

und wir erhalten weiterhin

$$w_s^{\text{smAIC}}(Y_n) = \frac{\exp(\text{AIC}(f_s | Y_n)/2)}{\sum_{s=1}^r \exp(\text{AIC}(f_s | Y_n)/2)} \xrightarrow{a.s.} \frac{\exp(a_s/2)}{\sum_{s=1}^r \exp(a_s/2)} =: w_s.$$

Insgesamt gilt dann analog zu Satz 4.1

$$\sqrt{n} \left(\sum_{s=1}^r w_s^{\text{smAIC}}(Y_n) \hat{\mu}_s - \sum_{s=1}^r w_s^{\text{smAIC}}(Y_n) \mu_s(\theta_s^*(\xi)) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_w^2(\theta^*(\xi))),$$

wobei

$$\sigma_w^2(\theta^*(\xi)) = \sum_{s=1}^r \sum_{t=1}^r w_s w_t \left(\frac{\partial \mu_s(\theta_s^*(\xi))}{\partial \theta_s} \right)^T A_s^{-1}(\theta_s^*(\xi)) B_{st}(\theta_s^*(\xi), \theta_t^*(\xi)) A_t^{-1}(\theta_t^*(\xi)) \frac{\partial \mu_t(\theta_t^*(\xi))}{\partial \theta_t}.$$

Wir betrachten nun ein Regressionsmodell der Form (2.1), in dem die Fehler ε_{ij} normalverteilt sind. Die wahre Dichte $g(\cdot | x_i, \theta)$ ist also die Dichte einer Normalverteilung mit Erwartungswertfunktion $\eta(x_i, \vartheta)$ und Varianz σ^2 . Die einzelnen Kandidatenmodelle hingegen besitzen die Dichten $f_s(\cdot | x_i, \theta_s)$ von Normalverteilungen mit Erwartungswertfunktion $\eta_s(x_i, \vartheta_s)$ und Varianz σ^2 . Mit dem Parameter $\theta_{s,g}^* = (\sigma_{s,*}^2, (\vartheta_s^*)^T)^T$ aus (2.14), der den Abstand zum wahren Parameterwert $\theta = (\sigma^2, \vartheta^T)^T$ minimiert, erhalten wir dann

$$\begin{aligned} & \mathbb{E}_{g(\cdot | x_i)} [\log f_s(Y_{ij} | x_i, \theta_{s,g}^*(\xi))] \\ &= -\log(\sqrt{2\pi}\sigma_{s,*}) - \frac{1}{2\sigma_{s,*}^2} \sum_{i=1}^k \xi_i \mathbb{E}_{g(\cdot | x_i)} (Y_{ij} - \eta_s(x_i, \vartheta_s^*))^2 \\ &= -\log(\sqrt{2\pi}\sigma_{s,*}) - \frac{\sigma^2}{2\sigma_{s,*}^2} - \frac{(\eta(x_i, \vartheta) - \eta_s(x_i, \vartheta_s^*))^2}{2\sigma_{s,*}^2}. \end{aligned}$$

Damit ergibt sich also

$$\begin{aligned} \text{AIC}(f_s | Y_n) &\xrightarrow{\text{a.s.}} -2\log(\sqrt{2\pi}\sigma_{s,*}) - \frac{\sigma^2}{\sigma_{s,*}^2} \\ &\quad - \sum_{i=1}^k \xi_i \frac{(\eta(x_i, \vartheta) - \eta_s(x_i, \vartheta_s^*))^2}{\sigma_{s,*}^2} - 2p_s. \end{aligned}$$

4.3 Anwendung und Simulation

In diesem Kapitel zeigen wir mit Hilfe einer Simulationsstudie, dass die Effizienz aller betrachteten Schätzmethode durch die Wahl eines geeigneten Versuchsplans deutlich gesteigert werden kann. Dazu betrachten wir erneut die Situation aus Abschnitt 4.1, in welchem wir Regressionsmodelle der Form (2.1) mit normalverteilten Fehlern mit Varianz σ^2 betrachtet haben. Wir betrachten zwei Mengen von Kandidatenmodellen, nämlich \mathcal{S}_1 aus (4.1), welche das log-lineare, das Emax und das quadratische Modell enthält, und \mathcal{S}_2 aus (4.2) mit dem log-linearen, dem Emax und dem exponentiellen Modell.

Mit dem Kriterium aus Definition 4.3 bestimmen wir im Folgenden Bayes-optimale Versuchspläne zur Model-Averaging Schätzung der $\text{ED}_{0.4}$ mit gleicher Gewichtung und $n = 100$ Beobachtungen. Dazu benötigen wir zunächst eine A-priori Verteilung für die unbekannte Dichte g , wozu wir eine Verteilung der Form (4.9) nutzen. Wir betrachten also die Dichten $f_s(y | x, \theta_s)$ von Normalverteilungen mit Erwartungswertfunktion $\eta_s(x, \vartheta_s)$ und Varianz $\sigma^2 = 0.1$, wobei die Erwartungswertfunktionen in Tabelle 4.1 gegeben sind. Für diese müssen die Mengen \mathcal{F}_{f_s} möglicher Parameterwerte festgelegt werden. Dabei ist zu beachten, dass das Kriterium (4.10) bei Schätzung der $\text{ED}_{0.4}$ nicht von dem Parameter ϑ_{s1} abhängt. Aus diesem Grund variieren wir diesen nicht, sondern wählen die Parameterspezifikationen aus Tabelle 4.1. Für die übrigen Parameter wählen wir jeweils den Wert aus Tabelle 4.1 sowie einen um 10% größeren beziehungsweise kleineren Wert dieses Parameters. Damit ergeben sich die Mengen

$$\begin{aligned} \mathcal{F}_{f_1} &= \{(0, \vartheta_{12}, \vartheta_{13}) : \vartheta_{12} = 0.0797 \pm 10\%, \vartheta_{13} = 1 \pm 10\%\}, \\ \mathcal{F}_{f_2} &= \{(0, \vartheta_{22}, \vartheta_{23}) : \vartheta_{22} = 0.467 \pm 10\%, \vartheta_{23} = 25 \pm 10\%\}, \\ \mathcal{F}_{f_3} &= \{(-0.08265, \vartheta_{32}, \vartheta_{33}) : \vartheta_{32} = 0.08265 \pm 10\%, \vartheta_{33} = 85 \pm 10\%\}, \\ \mathcal{F}_{f_4} &= \{(0, \vartheta_{42}, \vartheta_{43}) : \vartheta_{42} = 0.00533 \pm 10\%, \vartheta_{43} = -0.00002 \pm 10\%\}. \end{aligned} \tag{4.11}$$

4.3.1 Modelle von ähnlicher Form

Zunächst betrachten wir die Menge $\mathcal{S}_1 = \{f_1, f_2, f_4\}$ von Kandidatenmodellen, bestehend aus dem log-linearen, dem Emax und dem quadratischen Modell. Für diese Menge verwenden wir zur Konstruktion einer A-priori Verteilung (4.9) für das Kriterium (4.10) eine Gleichverteilung π_2 auf \mathcal{S}_1 sowie eine Gleichverteilung $\pi_1(\cdot | f_s)$ auf jeder Menge \mathcal{F}_{f_s} in (4.11) ($s = 1, 2, 4$). Der Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der $ED_{0.4}$ ergibt sich unter Verwendung des COBYLA-Algorithmus und ist gegeben durch

$$\xi_{\mathcal{S}_1}^* = \begin{Bmatrix} 0 & 14.447 & 49.283 & 150 \\ 0.227 & 0.167 & 0.365 & 0.240 \end{Bmatrix}. \quad (4.12)$$

Um die Effizienz dieses Versuchsplans in endlichen Stichproben zu untersuchen, vergleichen wir ihn mit dem Versuchsplan

$$\xi_P = \begin{Bmatrix} 0 & 10 & 25 & 50 & 100 & 150 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{Bmatrix}, \quad (4.13)$$

den Pinheiro et al. (2006) in einer Studie zur Dosis-Wirkungs-Beziehung von Beruhigungsmitteln verwenden (siehe auch Abschnitt 4.1) und dem lokal optimalen Versuchsplan zur Schätzung der $ED_{0.4}$ im log-linearen Modell f_1 mit Parameter $\vartheta_1 = (0, 0.0797, 1)^T$, welcher gegeben ist durch

$$\xi_L = \begin{Bmatrix} 0 & 4.051 & 150 \\ 0.339 & 0.5 & 0.161 \end{Bmatrix}. \quad (4.14)$$

Dieser Versuchsplan ist so konstruiert, dass er die asymptotische Varianz des Schätzers (2.12) der $ED_{0.4}$ im log-linearen Modell minimiert (siehe Dette et al. (2010)). Die Ergebnisse für die lokal optimalen Versuchspläne zur Schätzung der $ED_{0.4}$ im Emax und im exponentiellen Modell, welche Dette et al. (2010) ebenfalls herleiten, sind ähnlich und daher hier nicht angegeben. Wir beschränken uns zudem darauf, die Ergebnisse für $n = 100$ Beobachtungen zu präsentieren.

Wir führen mit obigen Versuchsplänen nun eine Simulationsstudie analog zu Abschnitt 4.1 durch. Insbesondere generieren wir in jedem Simulationsdurchlauf $n = 100$ Beobachtungen gemäß (4.3) unter Verwendung der Regressionsfunktionen und Parameter, die in Tabelle 4.1 gegeben sind. Wir verwenden dabei auch das Modell f_3 , welches nicht in der Menge \mathcal{S}_1 der Kandidatenmodelle enthalten ist, die zur Bestimmung des Model-Averaging Schätzers zur Verfügung stehen. Dabei nutzen wir jeweils einen der oben angegebenen Versuchspläne, wobei wir unter Umständen

Modell	Versuchsplan	Schätzmethode		
		Gleiche Gew.	Glatte AIC-Gew.	Modellselektion
f_1	$\xi_{\mathcal{S}_1}^*$	164.338	161.239	182.069
	ξ_P	223.291	218.99	285.062
	ξ_L	185.251	184.77	340.698
f_2	$\xi_{\mathcal{S}_1}^*$	122.665	135.969	168.577
	ξ_P	189.785	203.796	251.836
	ξ_L	501.814	501.394	1162.654
f_4	$\xi_{\mathcal{S}_1}^*$	174.535	212.746	331.552
	ξ_P	190.662	244.558	391.443
	ξ_L	404.716	427.548	1396.051
f_3	$\xi_{\mathcal{S}_1}^*$	1073.859	804.385	630.458
	ξ_P	1109.622	856.484	729.912
	ξ_L	3184.11	3413.566	4102.964

Tabelle 4.4: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.4}$ mit gleichen Gewichten (linke Spalte), mit glatten AIC-Gewichten (mittlere Spalte) und des Schätzers nach Modellselektion mit dem AIC (rechte Spalte). Die Menge der Kandidatenmodelle ist $\mathcal{S}_1 = \{f_1, f_2, f_4\}$. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen aus Tabelle 4.1 und verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan $\xi_{\mathcal{S}_1}^*$ zur Model-Averaging Schätzung der $ED_{0.4}$ aus (4.12). Zweite Zeile: Versuchsplan ξ_P aus (4.13) aus Pinheiro et al. (2006). Dritte Zeile: Lokal optimaler Versuchsplan ξ_L zur Schätzung der $ED_{0.4}$ aus (4.14).

die in Kapitel 2 beschriebene Rundungsprozedur aus Pukelsheim und Rieder (1992) nutzen, um die Anzahl n_i zu erhalten, mit der Versuche mit der Versuchsbedingung x_i wiederholt werden. Die gemäß (4.4) errechneten simulierten mittleren quadratischen Fehler finden sich in Tabelle 4.4. Dabei präsentieren die unterschiedlichen Spalten die Ergebnisse der unterschiedlichen Schätzmethode. In der linken Spalte finden sich die mittleren quadratischen Fehler des Model-Averaging Schätzers mit gleicher Gewichtung, während in der mittleren und rechten Spalte die Resultate des Model-Averaging Schätzers mit glatter AIC-Gewichtung beziehungsweise des Schätzers nach Modellselektion abgetragen sind. Die verschiedenen Zeilen zeigen jeweils die mittleren quadratischen Fehler unter Verwendung unterschiedlicher datengenerierender Modelle und unter Verwendung der verschiedenen Versuchspläne $\xi_{\mathcal{S}_1}^*$ aus (4.12), ξ_P aus (4.13) und ξ_L aus (4.14). Die fettgedruckten Zahlen markieren für

jede Schätzmethode den kleinsten mittleren quadratischen Fehler innerhalb der drei betrachteten Versuchspläne.

Wir diskutieren zunächst die Simulationsergebnisse unter Verwendung eines der in $\mathcal{S}_1 = \{f_1, f_2, f_4\}$ enthaltenen Modelle zur Generierung der Daten. Diese finden sich im oberen Teil von Tabelle 4.4. Wie auch in Abschnitt 4.1.1 ist zu beobachten, dass die beiden Model-Averaging Schätzer stets präziser sind als der Schätzer nach Modellselektion, wobei diese Beobachtung unabhängig vom betrachteten Versuchsplan ist. Im Vergleich zu den Versuchsplänen ξ_P und ξ_L erhöht der Bayes-optimale Versuchsplan $\xi_{\mathcal{S}_1}^*$ zur Model-Averaging Schätzung der $ED_{0.4}$ die Schätzgenauigkeit aller untersuchten Schätzmethoden. Betrachten wir beispielsweise das Szenario, in dem die Daten gemäß des log-linearen Modells f_1 generiert werden, so kann durch Verwendung des optimalen Versuchsplans der mittlere quadratische Fehler des Model-Averaging Schätzers mit festen Gewichten um 26.4% beziehungsweise 11.3% im Vergleich zu den Versuchsplänen ξ_P und ξ_L reduziert werden. Diese Verbesserung ist bemerkenswert, da der Versuchsplan ξ_L optimal ist zur Schätzung der $ED_{0.4}$ im log-linearen Modell f_1 , wenn die Daten auch gemäß dieses Modells generiert werden. Dieser Versuchsplan eignet sich somit offenbar nicht gut zur Schätzung der $ED_{0.4}$ im Emax oder im quadratischen Modell, welche ebenfalls im Model-Averaging Schätzer Berücksichtigung finden. Die höhere Effizienz, die durch Verwendung des optimalen Versuchsplans anstelle von ξ_P oder ξ_L resultiert, zeigt sich in den übrigen Szenarien noch deutlicher. So führt beispielsweise bei Nutzung des Emax Modells f_2 zur Erzeugung der Daten die Verwendung des optimalen Versuchsplans zu einer Verringerung des mittleren quadratischen Fehlers um 35.4% beziehungsweise 75.6% im Vergleich zu ξ_P und ξ_L . Tabelle 4.4 zeigt weiterhin, dass der optimale Versuchsplan (unabhängig vom datengenerierenden Modell) auch bei Verwendung des Model-Averaging Schätzers mit glatten AIC-Gewichten oder des Schätzers nach Modellselektion deutlich akkuratere Schätzergebnisse liefert als die Versuchspläne ξ_P und ξ_L . Entstammen die Daten beispielsweise dem log-linearen Modell f_1 , so kann der mittlere quadratische Fehler durch Verwendung von $\xi_{\mathcal{S}_1}^*$ anstelle von ξ_P beziehungsweise ξ_L um 26.4% beziehungsweise 12.7% für den Model-Averaging Schätzer mit glatten AIC-Gewichten sowie 36.1% beziehungsweise 46.6% für den Schätzer nach Modellselektion reduziert werden. Ähnliche Ergebnisse zeigen sich unter Verwendung von f_2 und f_4 .

Wir betrachten nun den Fall, dass die Daten gemäß des exponentiellen Modells f_3 erzeugt werden, welches nicht unter den Kandidatenmodellen in \mathcal{S}_1 ist. Selbst in diesem Szenario verbessert die Verwendung des Bayes-optimalen Versuchsplans die Schätz-

genauigkeit deutlich. Dies zeigt sich vor allem bei dem Model-Averaging Schätzer mit glatten AIC-Gewichten (Verringerung des mittleren quadratischen Fehlers um 6.1% beziehungsweise 76.4% im Vergleich zu ξ_P beziehungsweise ξ_L) und bei dem Schätzer nach Modellselektion (Verbesserung um 13.6% beziehungsweise 84.6%). Für den Model-Averaging Schätzer mit gleicher Gewichtung liegt diese Verbesserung dagegen bei 3.2% beziehungsweise 66.3%.

Zusammenfassend zeigen die Ergebnisse aus Tabelle 4.4, dass die Verwendung des Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung der $ED_{0.4}$ zu einer substantiellen Verringerung des mittleren quadratischen Fehler führen kann. Im Vergleich zu den Versuchsplänen ξ_P und ξ_L liegt die Verringerung für den Model-Averaging Schätzer mit gleicher Gewichtung bei 3.2%-75.6%, für den Model-Averaging Schätzer mit glatter AIC-Gewichtung bei 6.1%-76.4% und für den Schätzer nach Modellselektion bei 13.6%-85.5%.

4.3.2 Modelle von unterschiedlicher Form

Wir betrachten nun die zweite Menge $\mathcal{S}_2 = \{f_1, f_2, f_3\}$ von Kandidatenmodellen, welche aus dem log-linearen, dem Emax und dem exponentiellen Modell besteht. Hier verwenden wir zur Bestimmung einer A-priori Verteilung gemäß (4.9) eine Gleichverteilung π_2 auf der Menge \mathcal{S}_2 sowie eine Gleichverteilung $\pi_1(\cdot | f_s)$ auf jeder Menge \mathcal{F}_{f_s} in (4.11) ($s = 1, 2, 3$). Mit dieser A-priori Verteilung ist der numerisch bestimmte Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der $ED_{0.4}$ gegeben durch

$$\xi_{\mathcal{S}_2}^* = \left\{ \begin{array}{ccccc} 0 & 5.529 & 74.303 & 77.186 & 150 \\ 0.179 & 0.142 & 0.274 & 0.162 & 0.243 \end{array} \right\}. \quad (4.15)$$

Anders als der Versuchsplan $\xi_{\mathcal{S}_1}^*$ in (4.12) besitzt dieser fünf Trägerpunkte.

Auch mit diesem Versuchsplan führen wir eine Simulationsstudie analog zu Abschnitt 4.3.1 durch. Die Ergebnisse finden sich in Tabelle 4.5, wobei auch hier die unterschiedlichen Spalten die mittleren quadratischen Fehler der verschiedenen Schätzmethoden angeben und verschiedene Zeilen zu den unterschiedlichen Versuchsplänen $\xi_{\mathcal{S}_2}^*$ aus (4.15), ξ_P aus (4.13) und ξ_L aus (4.14) und den unterschiedlichen datengenerierenden Modellen f_1, f_2, f_3 und f_4 gehören.

In Tabelle 4.5 ist zu sehen, dass bei Verwendung des Bayes-optimalen Versuchsplans (anders als in Abschnitt 4.1.2) der Schätzer nach Modellselektion am präzisesten ist, sofern das datengenerierende Modell unter den Kandidatenmodellen ist. Die Verwendung von $f_4 \notin \mathcal{S}_2$ hingegen führt dazu, dass der Model-Averaging Schätzer

Modell	Versuchsplan	Schätzmethode		
		Gleiche Gew.	Glatte AIC-Gew.	Modellselektion
f_1	$\xi_{\mathcal{S}_2}^*$	724.682	349.615	344.991
	ξ_P	712.404	340.254	353.707
	ξ_L	798.473	444.457	459.335
f_2	$\xi_{\mathcal{S}_2}^*$	510.354	239.594	206.281
	ξ_P	517.963	267.967	286.272
	ξ_L	1155.793	1025.269	1835.355
f_3	$\xi_{\mathcal{S}_2}^*$	834.284	665.07	646.381
	ξ_P	871.362	766.140	802.763
	ξ_L	1526.230	1842.633	2721.415
f_4	$\xi_{\mathcal{S}_2}^*$	148.559	307.103	388.958
	ξ_P	208.628	298.315	419.651
	ξ_L	522.652	610.198	1907.066

Tabelle 4.5: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0.4}$ mit gleichen Gewichten (linke Spalte), mit glatten AIC-Gewichten (mittlere Spalte) und des Schätzers nach Modellselektion mit dem AIC (rechte Spalte). Die Menge der Kandidatenmodelle ist $\mathcal{S}_2 = \{f_1, f_2, f_3\}$. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen aus Tabelle 4.1 und verschiedenen Versuchsplänen. Erste Zeile: Bayes-optimaler Versuchsplan $\xi_{\mathcal{S}_2}^*$ zur Model-Averaging Schätzung der $ED_{0.4}$ aus (4.15). Zweite Zeile: Versuchsplan ξ_P aus (4.13) aus Pinheiro et al. (2006). Dritte Zeile: Lokal optimaler Versuchsplan ξ_L zur Schätzung der $ED_{0.4}$ aus (4.14).

mit gleichen Gewichten den kleinsten mittleren quadratischen Fehler aufweist. Weiterhin demonstriert Tabelle 4.5 auch für die Menge \mathcal{S}_2 von Kandidatenmodellen mit unterschiedlicher Form, dass in den meisten betrachteten Szenarien und für beinahe alle Schätzmethode die Verwendung des optimalen Versuchsplans die Schätzgenauigkeit im Vergleich zu den Versuchsplänen ξ_P und ξ_L erhöht. Es ist jedoch auch ersichtlich, dass unter Verwendung des log-linearen Modells in (4.3) der mittlere quadratische Fehler des Model-Averaging Schätzers mit gleichen Gewichten unter Verwendung des optimalen Versuchsplans um 1.7% größer ist als unter Verwendung von ξ_P . Für den Model-Averaging Schätzer mit glatten AIC-Gewichten beträgt diese Vergrößerung 2.8%. Dies ist damit zu begründen, dass der Bayes-optimale Versuchsplan so konstruiert ist, dass er für eine Menge von Modellen optimal ist. Für einzelne datengenerierende Modelle f_s kann er dadurch eine geringere Effizienz besitzen als andere Versuchspläne.

Zusammenfassend zeigen die numerischen Ergebnisse, dass der Bayes-optimale Versuchsplan $\xi_{\mathcal{S}_2}^*$ den mittleren quadratischen Fehler des Model-Averaging Schätzers mit gleichen Gewichten um bis zu 71.6% reduzieren kann im Vergleich zu den Versuchsplänen ξ_P und ξ_L . Diese Verbesserung kann für den Model-Averaging Schätzer mit glatten AIC-Gewichten und den Schätzer nach Modellselektion mit bis zu 76.6% beziehungsweise 88.8% noch größer ausfallen. Selbst für das quadratische Modell, welches sich nicht unter den Kandidatenmodellen befindet, können Verbesserungen beobachtet werden.

4.4 Ein adaptiver Ansatz

In den vorangegangenen Abschnitten haben wir gezeigt, dass mittels des Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung der $ED_{0.4}$ mit gleichen Gewichten oftmals eine deutliche Verbesserung der Schätzgenauigkeit des Model-Averaging Schätzers mit glatten AIC-Gewichten und auch des Schätzers nach Modellselektion erzielt werden kann, auch wenn der Versuchsplan unter der Annahme gleicher Modellgewichte konstruiert ist. Weiterhin zeigt sich insbesondere bei Verwendung der Menge \mathcal{S}_2 von Kandidatenmodellen, dass der Model-Averaging Schätzer mit datenbasierten Gewichten, sowie auch der Schätzer nach Modellselektion wesentlich genauer sein können, als der Model-Averaging Schätzer mit festen Gewichten. Daher ist es von Interesse zu untersuchen, ob diese beiden in der Praxis häufig genutzten Schätzmethoden zusätzlich verbessert werden können durch ein Versuchsplanungskriterium, das die Verwendung datenbasierter Gewichte besser reflektiert. Da jedoch, wie in Abschnitt 4.2 erläutert, der asymptotische mittlere quadratische Fehler eines Model-Averaging Schätzers mit zufälligen Gewichten im Allgemeinen nicht in geschlossener Form darstellbar ist, schlagen wir ein adaptives Vorgehen zur Verbesserung von Model-Averaging Schätzern mit zufälligen Gewichten vor, welches bereits in Abschnitt 3.4 in ähnlicher Form Anwendung gefunden hat. Wir beschränken uns auf zwei adaptive Schritte zur Bestimmung optimaler Versuchspläne, eine Erweiterung auf weitere Schritte ist jedoch möglich.

Wie auch in Abschnitt 3.4 teilen wir im Folgenden die Gesamtstichprobe mit n Beobachtungen in zwei Gruppen $n^{(1)}$ und $n^{(2)}$ auf. Im ersten Schritt erheben wir $n^{(1)}$ Beobachtungen gemäß des Bayes-optimalen Versuchsplans zur Model-Averaging Schätzung des Parameters μ mit gleichen Gewichten. Wir bezeichnen diese Stichprobe im Folgenden mit $Y_{n^{(1)}}^{(1)}$ und berechnen anhand dieser die glatten AIC-Gewichte $w_s^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$. Im zweiten Schritt der Versuchsplanung verwenden wir diese zur Be-

rechnung des mittleren quadratischen Fehlers in (4.7) und bestimmen den Bayes-optimalen Versuchsplan zur Model-Averaging Schätzung des Parameters μ mit Gewichten $w_1^{\text{smAIC}}(Y_{n^{(1)}}^{(1)}), \dots, w_r^{\text{smAIC}}(Y_{n^{(1)}}^{(1)})$. Diesen nutzen wir, um weitere $n^{(2)}$ Beobachtungen zu erheben. Bedingt auf die Stichprobe $Y_{n^{(1)}}^{(1)}$ ist der Model-Averaging Schätzer asymptotisch normalverteilt (für $n^{(2)} \rightarrow \infty$), sodass die Anwendung des Kriteriums (4.8) möglich ist. Abschließend nutzen wir die gesamte Stichprobe vom Umfang $n = n^{(1)} + n^{(2)}$, um den Zielparameter μ mittels eines Model-Averaging Schätzers mit glatten AIC-Gewichten oder mittels des Schätzers nach Modellselektion mit dem AIC zu schätzen.

Zur Bestimmung adaptiver Versuchspläne ist es notwendig, die Stichprobe geeignet aufzuteilen, da die Effizienz der resultierenden Versuchspläne maßgeblich von dieser Aufteilung abhängt. So zeigen beispielsweise Dette et al. (2013), dass die Effizienz adaptiver Versuchspläne stark von dem zugrundeliegenden statistischen Problem abhängt und erläutern, dass bei nichtlinearen Regressionsmodellen mit großen Varianzen die erste Stichprobe hinreichend groß gewählt werden sollte. Wir schlagen hier ein Vorgehen zur Aufteilung der Stichprobe vor, welches auf Simulationen beruht. Dazu betrachten wir zunächst die Situation aus Abschnitt 4.3. Hier erhöhen die nicht-adaptiven Bayes-optimalen Versuchspläne in beinahe allen betrachteten Situationen die Schätzgenauigkeit der Model-Averaging Schätzer. Bei Verwendung der Menge \mathcal{S}_2 von Kandidatenmodellen, welche aus dem log-linearen (f_1), dem Emax (f_2) und dem exponentiellen Modell (f_3) besteht, zeigt sich jedoch, dass der Model-Averaging Schätzer mit glatten AIC-Gewichten basierend auf dem Bayes-optimalen Versuchsplan einen größeren mittleren quadratischen Fehler besitzt, als der entsprechende Schätzer unter Verwendung des Versuchsplans ξ_P aus (4.13), falls das log-lineare Modell zur Erzeugung der Daten genutzt wird (siehe Tabelle 4.5). Wir nutzen dies als Startpunkt, um die „optimale“ Aufteilung der Stichprobe vom Umfang $n = 100$ zu bestimmen, da insbesondere in diesem Szenario eine Verbesserung im Vergleich zu ξ_P erreicht werden soll. Dazu betrachten wir verschiedene Stichprobenaufteilungen $n^{(1)}/n$ mittels derer wir den mittleren quadratischen Fehler der adaptiven Vorgehensweise simulieren. In insgesamt 1000 Simulationsdurchläufen bestimmen wir also jeweils einen Model-Averaging Schätzer mit glatten AIC-Gewichten, beziehungsweise einen Schätzer nach Modellselektion unter Verwendung der oben beschriebenen adaptiven Vorgehensweise für verschiedene $n^{(1)}$. Die Ergebnisse sind in Abbildung 4.2 zu finden. In dieser ist in der linken Grafik der resultierende mittlere quadratische Fehler des Model-Averaging Schätzers mit glatten AIC-Gewichten für unterschiedliche Anfangsstichprobengrößen $n^{(1)}$ abgetragen. Dabei entspricht $n^{(1)}/n = 1$ dem

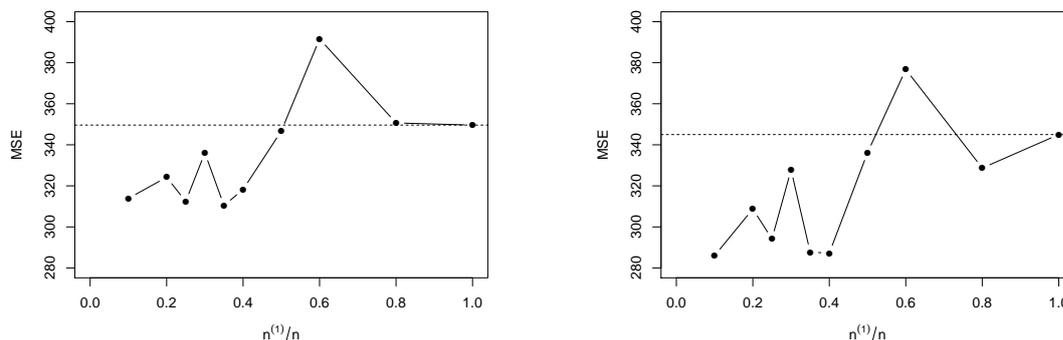


Abbildung 4.2: Mittlerer quadratischer Fehler des adaptiven Ansatzes für verschiedene Aufteilungen der Stichprobe. Links: Model-Averaging Schätzer mit glatten AIC-Gewichten. Rechts: Schätzer nach Modellselektion mit dem AIC.

mittleren quadratischen Fehler aus Tabelle 4.5, da hier de facto keine Anpassung der Gewichte stattfindet. In der rechten Grafik finden sich die entsprechenden Ergebnisse unter Verwendung des Schätzers nach Modellselektion mit dem AIC. Die Abbildungen verdeutlichen, dass durch Verwendung des adaptiven Ansatzes bis zu einer initialen Stichprobengröße von $n^{(1)} = 50$ der mittlere quadratische Fehler im Vergleich zum nicht-adaptiven Ansatz gesenkt werden kann. Für größere Stichprobenumfänge im ersten Schritt zeigt sich dagegen bei Verwendung des Model-Averaging Schätzers mit glatten AIC-Gewichten eine Verschlechterung der Schätzgenauigkeit. Bei Schätzung der $ED_{0.4}$ mittels Model-Averaging basierend auf glatten AIC-Gewichten ergibt sich der geringste mittlere quadratische Fehler für eine Anfangsstichprobengröße von $n^{(1)} = 35$. Auch für den Schätzer nach Modellselektion mit dem AIC liefert diese Stichprobenaufteilung gute Ergebnisse. Der resultierende mittlere quadratische Fehler von 287.699 ist lediglich um 0.6% größer als der minimale mittlere quadratische Fehler von 286.027 für $n^{(1)} = 10$ (siehe rechte Grafik von Abbildung 4.2). Daher wählen wir für alle folgenden Simulationen $n^{(1)} = 35$ für die erste Stichprobe und $n^{(2)} = 65$ für die zweite Stichprobe.

Wir betrachten nun die Situation aus Abschnitt 4.3.1 und Abschnitt 4.3.2, um mittels Simulationen die Effizienz des oben beschriebenen adaptiven Verfahrens mit der Stichprobenaufteilung $n^{(1)}/100 = 0.35$ im Vergleich zu den in Abschnitt 4.3 betrachteten Versuchsplänen zu untersuchen. Dazu gibt Tabelle 4.6 die entsprechenden Ergebnisse der Simulationen an. Für die Situation aus Abschnitt 4.3.1, bei der die Kandidatenmodelle der Menge \mathcal{S}_1 jeweils eine ähnliche Form haben, findet sich im linken Teil der Tabelle der mittlere quadratische Fehler des Model-Averaging Schät-

Modell	Kandidatenmodelle: $\{f_1, f_2, f_4\}$		Kandidatenmodelle: $\{f_1, f_2, f_3\}$	
	glatte AIC-Gew.	Modellselektion	glatte AIC-Gew.	Modellselektion
f_1	184.500	198.329	310.433	287.699
f_2	135.033	168.504	258.495	228.569
f_3	783.574	623.298	698.500	645.918
f_4	225.692	349.164	253.726	306.753

Tabelle 4.6: Mittlerer quadratischer Fehler verschiedener Schätzer der $ED_{0,4}$ unter Verwendung des adaptiven Vorgehens. Die Menge der Kandidatenmodelle ist $\mathcal{S}_1 = \{f_1, f_2, f_4\}$ (linker Teil) und $\mathcal{S}_2 = \{f_1, f_2, f_3\}$ (rechter Teil). Linke Spalten: Model-Averaging Schätzer mit glatten AIC-Gewichten; Rechte Spalten: Schätzer nach Modellselektion mit dem AIC. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen.

zers mit glatten AIC-Gewichten und des Schätzers nach Modellselektion mit dem AIC. Dabei ist zu beachten, dass f_3 nicht in der Menge \mathcal{S}_1 enthalten ist. Wie bereits in Abschnitt 4.1.1 ist auch hier der Model-Averaging Schätzer mit glatten AIC Gewichten präziser als der Schätzer nach Modellselektion, sofern das datengenerierende Modell unter den Kandidatenmodellen ist. Die Ergebnisse können jeweils mit den fettgedruckten Ergebnissen aus Tabelle 4.4 verglichen werden, welche dem mittleren quadratischen Fehler der verschiedenen Schätzer basierend auf dem nicht-adaptiven Bayes-optimalen Versuchsplan $\xi_{\mathcal{S}_1}^*$ aus (4.12) entsprechen. Der Vergleich der mittleren quadratischen Fehler offenbart, dass das adaptive Vorgehen bessere Ergebnisse liefert als das nicht-adaptive Vorgehen, wenn das Emax, das quadratische oder das exponentielle Modell die Daten generieren. Die Nutzung des log-linearen Modells hingegen bewirkt, dass der Versuchsplan $\xi_{\mathcal{S}_1}^*$ geringere mittlere quadratische Fehler aufweist als das adaptive Vorgehen. Dennoch ist das adaptive Vorgehen effizienter als die zum Vergleich hinzugezogenen Versuchspläne ξ_P und ξ_L (siehe Tabelle 4.4). Im rechten Teil von Tabelle 4.6 finden sich die entsprechenden Ergebnisse für die Situation aus Abschnitt 4.3.2, in welchem wir die Menge von Kandidatenmodellen $\mathcal{S}_2 = \{f_1, f_2, f_3\}$ betrachtet haben, die mit dem log-linearen, dem Emax und dem exponentiellen Modell sehr unterschiedliche Kandidatenmodelle enthält. Hier weist der Model-Averaging Schätzer mit glatten AIC-Gewichten größere mittlere quadratische Fehler auf als der Schätzer nach Modellselektion mit dem AIC, wenn eines der in \mathcal{S}_2 enthaltenen Modelle zu Erzeugung der Daten genutzt wird. Um zu prüfen, ob das adaptive Vorgehen eine Verbesserung im Vergleich zu den in Abschnitt 4.3.2 betrachteten Versuchsplänen bietet, können die Resultate mit den in Tabelle 4.5 angegebenen mittleren quadratischen Fehlern verglichen werden. Zunächst wird deutlich,

dass bei Verwendung des log-linearen Modells f_1 in der Simulation der mittlere quadratische Fehler des Model-Averaging Schätzers mit glatten AIC-Gewichten durch Verwendung des adaptiven Vorgehens anstelle des optimalen Versuchsplans $\xi_{\mathcal{S}_2}^*$ um 11.2% reduziert werden kann. Insbesondere ist das adaptive Vorgehen zudem effizienter als die Verwendung des Versuchsplans ξ_P , welcher in Tabelle 4.5 den geringsten mittleren quadratischen Fehler aufweist. Bei Betrachtung des Schätzers nach Modellselektion liegt die Reduktion im Vergleich zu $\xi_{\mathcal{S}_2}^*$ bei 16.6%. Es ist zu beachten, dass die Aufteilung der Stichprobe so erfolgt ist, dass sie in dieser Situation „optimal“ ist. Diese Aufteilung muss jedoch nicht unbedingt geeignet sein, wenn andere Modelle die Daten erzeugen. So zeigt sich für das adaptive Vorgehen sowohl für das Emax Modell f_2 als auch das exponentielle Modell f_3 ein größerer mittlerer quadratischer Fehler als unter Verwendung des Bayes-optimalen Versuchsplans $\xi_{\mathcal{S}_2}^*$. Dennoch ist das hier vorgeschlagene adaptive Vorgehen stets besser als die Versuchspläne ξ_P aus (4.13) und ξ_L aus (4.14).

4.5 Approximation des Versuchsplanungskriteriums

Die Bestimmung von lokal oder Bayes-optimalen Versuchsplänen zur Model-Averaging Schätzung des Parameters μ ist sehr aufwändig. Sie muss numerisch erfolgen, wobei der COBYLA-Algorithmus zum Einsatz kommt. Dieser erfordert zahlreiche Berechnungen der Funktion $\Phi_{\text{mav}}^\pi(\xi)$ aus (4.8), wobei für jede Kalkulation zunächst der Parameterwert $\theta_{s,g}^*(\xi)$ aus (2.14) bestimmt werden muss. In diesem werten wir die Gradienten der Zielfunktion sowie die Matrizen A_s aus (2.17) und B_{st} aus (2.18) aus. Hierdurch müssen zahlreiche weitere Optimierungsschritte in jeder Iteration des COBYLA-Algorithmus durchgeführt werden und die Bestimmung Bayes-optimaler Versuchspläne gemäß Definition 4.3 nimmt mitunter viel Zeit in Anspruch.

Wir schlagen daher als Alternative zur Verwendung des exakten asymptotischen mittleren quadratischen Fehlers aus (4.7) eine Modifikation dessen vor. Wir ersetzen dabei den Parameter $\theta_{s,g}^*(\xi)$ aus (2.14) durch eine Näherung, welche nicht vom gewählten Versuchsplan abhängt. Genauer gesagt verwenden wir nun anstelle des Maßes ξ zur Berechnung des Kullback-Leibler-Abstands aus (2.13) das Lebesgue-Maß λ und definieren

$$\widetilde{\text{KL}}(g : f_s | \theta_s) = \int \int g(y | x) \log \left(\frac{g(y | x)}{f_s(y | x, \theta_s)} \right) dy d\lambda(x).$$

Damit ist der Parameter $\tilde{\theta}_{s,g}^*$, welcher

$$\tilde{\theta}_{s,g}^* = \arg \min_{\theta_s \in \Theta_s} \widetilde{\text{KL}}(g : f_s \mid \theta_s) \quad (4.16)$$

erfüllt, unabhängig vom Versuchsplan ξ . Wir verwenden nun diesen Parameter zur Berechnung des mittleren quadratischen Fehlers in (4.7) und erhalten mit

$$\tilde{\theta}_g^* = \left(\left(\tilde{\theta}_{1,g}^* \right)^T, \dots, \left(\tilde{\theta}_{r,g}^* \right)^T \right)^T$$

eine Approximation des mittleren quadratischen Fehlers der asymptotischen Verteilung des Model-Averaging Schätzers $\hat{\mu}_{\text{mav}}$

$$\tilde{\Phi}_{\text{mav}}(\xi, g) = \frac{1}{n} \sigma_w^2(\tilde{\theta}_g^*) + \left(\sum_{s=1}^r w_s \mu_s(\tilde{\theta}_{s,g}^*) - \mu_{\text{true}} \right)^2. \quad (4.17)$$

In diesem Ausdruck ist nun die Verzerrung unabhängig vom gewählten Versuchsplan ξ , sodass eine Minimierung des approximativen mittleren quadratischen Fehlers einer Minimierung der approximativen Varianz entspricht.

Definition 4.5. Unter den Annahmen aus Definition 4.2 heißt ein Versuchsplan ξ^* *approximativ lokal optimal zur Model-Averaging Schätzung des Parameters μ* , falls er die Funktion $\tilde{\Phi}_{\text{mav}}(\xi, g)$ aus (4.17) unter allen approximativen Versuchsplänen ξ auf \mathcal{X} minimiert, das heißt falls

$$\tilde{\Phi}_{\text{mav}}(\xi^*, g) = \min_{\xi \in \Xi} \tilde{\Phi}_{\text{mav}}(\xi, g).$$

Weiterhin heißt für eine A-priori Verteilung π auf \mathcal{G} der Versuchsplan ξ^* *approximativ Bayes-optimal zur Model-Averaging Schätzung des Parameters μ hinsichtlich der A-priori Verteilung π* , falls er die Funktion

$$\tilde{\Phi}_{\text{mav}}^\pi(\xi) = \int_{\mathcal{G}} \tilde{\Phi}_{\text{mav}}(\xi, g) d\pi(g) \quad (4.18)$$

minimiert, das heißt, falls

$$\tilde{\Phi}_{\text{mav}}^\pi(\xi^*) = \min_{\xi \in \Xi} \tilde{\Phi}_{\text{mav}}^\pi(\xi).$$

Wir gehen dabei stets davon aus, dass das Integral in (4.18) existiert.

Der Ansatz aus Definition 4.5 ist nicht nur numerisch weniger fordernd als das in 4.2 hergeleitete Kriterium, sondern wir können mit Satz 2.5 auch eine notwendige Bedingung herleiten, um zu überprüfen ob ein numerisch bestimmter Versuchsplan optimal sein kann.

Satz 4.6. *Ist ein Versuchsplan ξ^* Bayes-optimal zur Model-Averaging Schätzung des Parameters μ hinsichtlich der A-priori-Verteilung π , dann gilt für alle $x \in \mathcal{X}$*

$$d_\pi(x, \xi^*) = - \int_G \sum_{s=1}^r \sum_{t=1}^r w_s w_t \left(\frac{\partial \mu_s(\tilde{\theta}_{s,g}^*)}{\partial \theta_s} \right)^T h(\xi^*, x, \tilde{\theta}_{s,g}^*, \tilde{\theta}_{t,g}^*) \frac{\partial \mu_t(\tilde{\theta}_{t,g}^*)}{\partial \theta_t} d\pi \leq 0. \quad (4.19)$$

Dabei ist

$$\begin{aligned} h(\xi, x, \theta_s, \theta_t) = & A_s^{-1}(\theta_s, \xi) \cdot (B_{st}(\theta_s, \theta_t, \xi) + B_{st}(\theta_s, \theta_t, \xi_x) \\ & - A_s(\theta_s, \xi_x) A_s^{-1}(\theta_s, \xi) B_{st}(\theta_s, \theta_t, \xi) \\ & - B_{st}(\theta_s, \theta_t, \xi) A_t^{-1}(\theta_t, \xi) A_t(\theta_t, \xi_x)) A_t^{-1}(\theta_t, \xi), \end{aligned}$$

wobei ξ_x das in (2.6) definierte Einpunktdesign ist.

Darüber hinaus gilt die Gleichheit in (4.19) für alle Trägerpunkte x von ξ^* .

Bemerkung. Mit Satz 4.6 ergibt sich ebenfalls eine notwendige Bedingung für lokal optimale Versuchspläne zur Model-Averaging Schätzung des Parameters μ gemäß Definition 4.5. Hierzu ist lediglich eine Einpunktverteilung als A-priori Verteilung zu wählen.

Wir zeigen nun exemplarisch für die Situation aus Abschnitt 4.3.2, wie sich ein gemäß Definition 4.5 konstruierter Versuchsplan in Simulationen verhält. Wir betrachten also die Menge $\mathcal{S}_2 = \{f_1, f_2, f_3\}$ und bestimmen einen approximativ Bayes-optimalen Versuchsplan zur Model-Averaging Schätzung der $ED_{0.4}$ mit gleichen Gewichten. Dabei verwenden wir die in Abschnitt 4.3.2 erläuterte A-priori Verteilung und gehen von $n = 100$ Beobachtungen aus. Minimierung des Ausdrucks (4.18) liefert dann

$$\tilde{\xi}_{\mathcal{S}_2} = \left\{ \begin{array}{cccc} 0 & 9.801 & 78.982 & 150 \\ 0.169 & 0.169 & 0.447 & 0.215 \end{array} \right\}. \quad (4.20)$$

Dieser unterscheidet sich von $\xi_{\mathcal{S}_2}^*$ aus (4.15) vor allem darin, dass er lediglich vier Trägerpunkte besitzt. Für diesen Versuchsplan können wir die notwendige Bedingung aus Satz 4.6 berechnen. Diese ist in Abbildung 4.3 dargestellt und spricht nicht gegen das Vorliegen eines optimalen Versuchsplans.

Um zu untersuchen, wie sich der approximative Versuchsplan im Vergleich zum Bayes-optimalen Versuchsplan $\xi_{\mathcal{S}_2}^*$ aus (4.15) verhält, führen wir eine Simulation analog zu Abschnitt 4.3 durch. Die simulierten mittleren quadratischen Fehler, die sich bei Verwendung des Versuchsplans $\tilde{\xi}_{\mathcal{S}_2}$ aus (4.20) für verschiedene datengenerierende Modelle ergeben, finden sich in Tabelle 4.7. Die verschiedenen Spalten zeigen dabei die Ergebnisse für die verschiedenen Schätzmethoden.

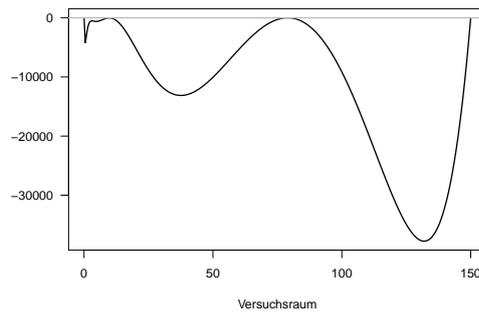


Abbildung 4.3: Die Funktion d_π aus (4.19) ausgewertet für den Versuchsplan $\tilde{\xi}_{\mathcal{S}_2}$ aus (4.20).

Wir konzentrieren uns zunächst auf die mittleren quadratischen Fehler, die sich unter Verwendung eines der datengenerierenden Modelle ergeben, die in $\mathcal{S}_2 = \{f_1, f_2, f_3\}$ enthalten sind. Hier zeigt sich, dass der approximative Versuchsplan für Daten, die gemäß des log-linearen Modells f_1 simuliert werden, den mittleren quadratischen Fehler aller Schätzmethode im Vergleich zu dem Bayes-optimalen Versuchsplan $\xi_{\mathcal{S}_2}^*$ senkt. So reduziert sich der mittlere quadratische Fehler um 5.3% (Model-Averaging Schätzer mit gleichen Gewichten), 10.9% (Model-Averaging Schätzer mit glatten AIC-Gewichten) beziehungsweise 13.7% (Schätzer nach Modellselektion). Es ist jedoch zu beachten, dass der adaptive Ansatz aus Abschnitt 4.4 bessere Resultate aufweist als der approximative Versuchsplan (vergleiche Tabelle 4.6). Sind dagegen die Daten gemäß des Emax (f_2) oder des exponentiellen Modells (f_3) entstanden, so verringert sich die Schätzgenauigkeit durch Verwendung des approximativen Versuchsplans $\tilde{\xi}_{\mathcal{S}_2}$. Genauer gesagt erhöht sich der mittlere quadratische Fehler durch Verwendung von $\tilde{\xi}_{\mathcal{S}_2}$ anstelle von $\xi_{\mathcal{S}_2}^*$ bei Nutzung von f_2 beziehungsweise f_3 um 5.3% beziehungsweise 5.5% (Model-Averaging Schätzer mit gleichen Gewichten), 19.0% beziehungsweise 12.3% (Model-Averaging Schätzer mit glatten AIC-Gewichten) und 31.2% beziehungsweise 6.6% (Schätzer nach Modellselektion).

Betrachten wir die Ergebnisse für das quadratische Modell, welches bei der Bestimmung beider Versuchspläne keine Berücksichtigung findet, so ist für feste Gewichte der Bayes-optimale Versuchsplan effizienter. Die verschiedenen Schätzer basierend auf dem AIC hingegen sind präziser unter Verwendung des approximativen Versuchsplans.

Zusammenfassend weist der approximative Versuchsplan teils einen kleineren mittleren quadratischen Fehler auf als der Bayes-optimale Versuchsplan zur Model-Averaging Schätzung der $ED_{0.4}$. Auf der anderen Seite verschlechtert sich die Schätz-

Modell	Schätzmethode		
	Gleiche Gew.	Glatte AIC-Gew.	Modellselektion
f_1	686.479	311.352	297.852
f_2	537.628	285.067	270.580
f_3	880.255	747.186	689.152
f_4	159.123	275.181	336.636

Tabelle 4.7: Mittlerer quadratischer Fehler des Model-Averaging Schätzers der $ED_{0,4}$ mit gleichen Gewichten (linke Spalte), mit glatten AIC-Gewichten (mittlere Spalte) und des Schätzers nach Modellselektion mit dem AIC (rechte Spalte). Die Menge der Kandidatenmodelle ist $\mathcal{S}_2 = \{f_1, f_2, f_3\}$. Die verschiedenen Zeilen gehören zu verschiedenen datengenerierenden Modellen aus Tabelle 4.1 unter Verwendung des Versuchsplans $\tilde{\xi}_{\mathcal{S}_2}$ aus (4.20).

genauigkeit in einigen Situationen durch Verwendung dieses Versuchsplans jedoch deutlich. Dennoch kann dieser Ansatz bei einer großen Menge an Kandidatenmodellen hilfreich sein, da die Bestimmung Bayes-optimaler Versuchspläne gemäß Definition 4.3 sehr aufwändig ist und keine Überprüfung der numerisch bestimmten Versuchspläne möglich ist.

4.6 Beweise

Satz 4.1. Gleichung (A.2) aus White (1982) liefert zunächst

$$\sqrt{n}(\hat{\theta}_{n,s} - \theta_s^*(\xi)) + A_s^{-1}(\theta_s^*(\xi)) \frac{1}{\sqrt{n}} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{\partial \log f_s(Y_{ij} | x_i, \theta_s^*(\xi))}{\partial \theta_s} \xrightarrow{p} 0, \quad (4.21)$$

wobei \xrightarrow{p} Konvergenz in Wahrscheinlichkeit bezeichnet. Dabei ist die Matrix $A_s(\theta_s^*) = A_s(\theta_s^*(\xi), \xi)$ aus (2.17) nicht singulär dank Annahme (A6). Durch Anwendung des multivariaten zentralen Grenzwertsatzes (siehe beispielsweise (Rao, 1973, Kapitel 2)) erhalten wir

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{\partial \log f_1(Y_{ij}|x_i, \theta_1^*(\xi))}{\partial \theta_1} \\ \vdots \\ \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{\partial \log f_r(Y_{ij}|x_i, \theta_r^*(\xi))}{\partial \theta_r} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \begin{pmatrix} B_{11} & \dots & B_{1r} \\ \vdots & \ddots & \vdots \\ B_{r1} & \dots & B_{rr} \end{pmatrix} \right), \quad (4.22)$$

wobei $B_{st} = B_{st}(\theta_s^*(\xi), \theta_t^*(\xi), \xi)$ in (2.18) definiert ist.

Durch Kombination von (4.21) und (4.22) ergibt sich die schwache Konvergenz des Vektors $\hat{\theta}_n = (\hat{\theta}_{n,1}^T, \dots, \hat{\theta}_{n,r}^T)^T$, nämlich

$$\sqrt{n}(\hat{\theta}_n - \theta^*(\xi)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

wobei $\Sigma = (\Sigma_{st})_{s,t=1,\dots,r}$ eine Blockmatrix mit $p_s \times p_t$ Einträgen

$$\Sigma_{st} = A_s^{-1}(\theta_s^*(\xi)) B_{st}(\theta_s^*(\xi), \theta_t^*(\xi)) A_t^{-1}(\theta_t^*(\xi)), \quad s, t = 1, \dots, r$$

ist und der Vektor $\theta^*(\xi)$ gegeben ist durch

$$\theta^*(\xi) = ((\theta_1^*(\xi))^T, \dots, (\theta_r^*(\xi))^T)^T.$$

Weiterhin definieren wir für den Parametervektor $\tilde{\theta}^T = (\theta_1^T, \dots, \theta_r^T) \in \mathbb{R}^{\sum_{s=1}^r p_s}$ die Projektion π_s durch $\pi_s \tilde{\theta} := \theta_s$ und den Vektor

$$\tilde{\mu}(\tilde{\theta}) = (\mu_1(\pi_1 \tilde{\theta}), \dots, \mu_r(\pi_r \tilde{\theta}))^T$$

mit Ableitung

$$\mu'_{\tilde{\theta}} = \begin{pmatrix} \left(\frac{\partial \mu_1(\theta_1)}{\partial \theta_1} \right)^T & 0 & \dots & 0 \\ 0 & \left(\frac{\partial \mu_2(\theta_2)}{\partial \theta_2} \right)^T & 0 & \dots & 0 \\ 0 & \dots & 0 & \left(\frac{\partial \mu_r(\theta_r)}{\partial \theta_r} \right)^T \end{pmatrix}.$$

Unter Verwendung der Delta-Methode (siehe beispielsweise Van der Vaart (2007, Kapitel 3)) resultiert

$$\sqrt{n}(\tilde{\mu}(\hat{\theta}_n) - \tilde{\mu}(\theta^*(\xi))) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mu'_{\theta^*(\xi)} \Sigma (\mu'_{\theta^*(\xi)})^T\right).$$

Die Aussage folgt schließlich durch das *Continuous Mapping Theorem* und die Darstellung

$$\hat{\mu}_{\text{mav}} = (w_1, \dots, w_r) \tilde{\mu}(\hat{\theta}_n).$$

□

Satz 4.6. Auch hier nutzen wir die Argumente aus Pukelsheim (1993, Kapitel 11) und gehen davon aus, dass Integration und Differenziation vertauschbar sind. Dann muss ein approximativ Bayes-optimaler Versuchsplan zur Model-Averaging Schätzung des Parameters μ nach Satz 2.5 die Ungleichheit

$$-D\tilde{\Phi}_{\text{mav}}^{\pi}(\xi^*)(\xi_x - \xi^*) = - \int_{\mathcal{G}} D\tilde{\Phi}_{\text{mav}}(\xi^*, g)(\xi_x - \xi^*) d\pi \leq 0 \quad (4.23)$$

für alle $x \in \mathcal{X}$ erfüllen. Dabei ist $D\tilde{\Phi}_{\text{mav}}(\xi^*, g)(\xi_x - \xi^*)$ die Richtungsableitung (2.4) der Funktion $\tilde{\Phi}_{\text{mav}}$ aus (4.17) in ξ^* in Richtung $\xi_x - \xi^*$ und ξ_x das Dirac-Maß im Punkt $x \in \mathcal{X}$ aus (2.6). Es sei darauf hingewiesen, dass das Versuchsplanungskriterium (4.8) nicht konvex ist und somit die Bedingung (4.23) nicht hinreichend ist.

Unter Berücksichtigung der Tatsache, dass (4.18) von ξ lediglich über $\sigma_w^2(\theta)$ aus (4.6) abhängt, genügt es die Richtungsableitung ebendieser Funktion in ξ^* in Richtung $\xi_x - \xi^*$ zu bestimmen. Dazu nutzen wir, dass die Richtungsableitung der Matrix $A_s^{-1}(\theta_s, \xi)$ in ξ^* in Richtung $\xi_x - \xi^*$ gegeben ist durch

$$DA_s^{-1}(\theta_s, \xi^*)(\xi^* - \xi_x) = A_s^{-1}(\theta_s, \xi^*) - A_s^{-1}(\theta_s, \xi^*)A_s(\theta_s, \xi_x)A_s^{-1}(\theta_s, \xi^*).$$

Unter Anwendung der Kettenregel erhalten wir dann für die Sandwich-Matrix in $\sigma_w^2(\theta)$ die Richtungsableitung

$$\begin{aligned} & DA_s^{-1}(\theta_s, \xi^*)B_{st}(\theta_s, \theta_t, \xi^*)A_t^{-1}(\theta_t, \xi^*)(\xi^* - \xi_x) \\ &= A_s^{-1}(\theta_s, \xi^*) \cdot (B_{st}(\theta_s, \theta_t, \xi^*) + B_{st}(\theta_s, \theta_t, \xi_x) - A_s(\theta_s, \xi_x)A_s^{-1}(\theta_s, \xi^*)B_{st}(\theta_s, \theta_t, \xi^*) \\ &\quad - B_{st}(\theta_s, \theta_t, \xi^*)A_t^{-1}(\theta_t, \xi^*))A_t(\theta_t, \xi_x))A_t^{-1}(\theta_t, \xi^*). \end{aligned}$$

Setzen wir nun $\tilde{\theta}_{s,g}^*$ aus (4.16) ein, so folgt die notwendige Bedingung (4.19).

Abschließend zeigen wir, dass in (4.19) Gleichheit für alle Trägerpunkte x des Versuchsplans ξ^* gilt. Dazu gehen wir zunächst davon aus, es gäbe einen Trägerpunkt x_0 von ξ , sodass wir echte Ungleichheit in (4.19) haben. Dann ergibt sich aber auch

$$- \int_{\mathcal{X}} \int_{\mathcal{G}} \sum_{s=1}^r \sum_{t=1}^r w_s w_t \left(\frac{\partial \mu_s(\tilde{\theta}_{s,g}^*)}{\partial \theta_s} \right)^T h(\xi^*, x, \tilde{\theta}_{s,g}^*, \tilde{\theta}_{t,g}^*) \frac{\partial \mu_t(\tilde{\theta}_{t,g}^*)}{\partial \theta_t} d\pi \xi^*(dx) < 0. \quad (4.24)$$

Da aber $\int_{\mathcal{X}} A_s(\theta_s, \xi_x) \xi^*(dx) = A_s(\theta_s, \xi^*)$ und $\int_{\mathcal{X}} B_{st}(\theta_s, \theta_t, \xi_x) \xi^*(dx) = B_{st}(\theta_s, \theta_t, \xi^*)$ gilt, ergibt sich andererseits

$$\int_{\mathcal{X}} \int_{\mathcal{G}} h(\xi^*, x, \tilde{\theta}_{s,g}^*, \tilde{\theta}_{t,g}^*) = 0 \quad (s, t = 1, \dots, r)$$

und damit

$$- \int_{\mathcal{X}} \int_{\mathcal{G}} \sum_{s=1}^r \sum_{t=1}^r w_s w_t \left(\frac{\partial \mu_s(\tilde{\theta}_{s,g}^*)}{\partial \theta_s} \right)^T h(\xi^*, x, \tilde{\theta}_{s,g}^*, \tilde{\theta}_{t,g}^*) \frac{\partial \mu_t(\tilde{\theta}_{t,g}^*)}{\partial \theta_t} d\pi \xi^*(dx) = 0.$$

Dies steht im Widerspruch zu (4.24). Folglich muss in (4.19) die Gleichheit gelten, wenn x ein Trägerpunkt des Versuchsplans ξ^* ist. \square

Kapitel 5

Zusammenfassung und Ausblick

Wir bestimmen in dieser Arbeit optimale Versuchspläne zur Model-Averaging Schätzung eines Zielparameters μ . Um uns diesem Ziel zu nähern, beschränken wir uns zunächst auf einen Spezialfall, bei dem wir annehmen, dass sämtliche Modelle verschachtelt sind und die Abweichungen von den wahren Parametern mit der Stichprobengröße sinken. Hierdurch können wir die Theorie aus Hjort und Claeskens (2003) nutzen, um handliche Ausdrücke für die asymptotische Verteilung von Model-Averaging Schätzern zu erhalten, bei denen Varianz und Verzerrung des Schätzers jeweils gleichermaßen mit n sinken. Wir schlagen basierend hierauf ein neuartiges Versuchsplanungskriterium vor, welches den asymptotischen mittleren quadratischen Fehler des Schätzers minimiert. Anhand einer Simulationsstudie zeigen wir, dass diese optimalen Versuchspläne die Effizienz von Model-Averaging Schätzern stark erhöhen können. Wir erweitern anschließend diese Theorie auf allgemeinere Situationen, in denen wir keine Annahmen an die Zusammenhänge zwischen den Kandidatenmodellen und der wahren Verteilung machen. Insbesondere kann mit dieser Theorie auch das Verhalten von Model-Averaging Schätzern bei nicht-verschachtelten Modellen erklärt werden. Auch für diese Situation leiten wir ein Versuchsplanungskriterium her, das den mittleren quadratischen Fehler des Model-Averaging Schätzers minimiert. In Simulationen zeigen auch die so bestimmten Versuchspläne eine höhere Schätzgenauigkeit (bezüglich des mittleren quadratischen Fehlers) als andere Versuchspläne.

Die oben beschriebenen Kriterien beruhen jeweils auf der Annahme, dass die Gewichte der Model-Averaging Schätzer fest gewählt sind. Dies ist in der Praxis für gewöhnlich jedoch nicht der Fall. Stattdessen werden oftmals Gewichte basierend auf Informationskriterien wie dem AIC oder dem BIC genutzt oder auch die Schätzer nach Modellselektion mit diesen Kriterien. Auch wenn die Simulationen zeigen, dass

die optimalen Versuchspläne für feste Modellgewichte den mittleren quadratischen Fehler von Schätzern dieser Art oftmals ebenfalls verbessern, so bleibt es eine herausfordernde Aufgabe ein Versuchsplanungskriterium zu entwickeln, dass diese Art von Modellgewichten angemessen berücksichtigt.

Es ist nicht in allen Situationen möglich, den mittleren quadratischen Fehler von Model-Averaging Schätzern mit datenabhängigen Gewichten in geschlossener Form anzugeben, sodass unter Umständen auf rechenintensive Simulationen oder numerische Integration zurückgegriffen werden muss, um diesen zu bestimmen. Die direkte Einbindung von zufälligen Modellgewichten in die Versuchsplanung ist daher eine herausfordernde Aufgabe in zukünftigen Studien. Als ersten Ansatz hierzu schlagen wir in dieser Arbeit einen adaptiven Ansatz vor, bei dem in einem ersten Schritt Beobachtungen gemäß des optimalen Versuchsplans zur Model-Averaging Schätzung mit gleichen Modellgewichten erhoben werden. Mittels dieser werden neue Modellgewichte bestimmt, welche zur Konstruktion eines neuen Versuchsplans genutzt werden. In Simulationen zeigt sich, dass in einigen Situationen die Schätzgenauigkeit durch Verwendung dieses Verfahrens verbessert werden kann. Oftmals liefert jedoch die Verwendung des optimalen Versuchsplans mit festen Modellgewichten geringere mittlere quadratische Fehler als das adaptive Vorgehen. Dies kann unter anderem an einer nicht optimalen Aufteilung der Gesamtstichprobe liegen. In Kapitel 4 schlagen wir für diese Aufteilung eine Vorgehensweise basierend auf Simulationen vor, die jedoch nur eines der Kandidatenmodelle berücksichtigt. Solch eine optimale Aufteilung der Stichprobe kann zu deutlichen Verbesserungen führen, wenn die Daten gemäß dieses spezifischen Kandidatenmodells erzeugt werden. Unter Verwendung anderer datengenerierender Modelle kann diese Aufteilung jedoch auch zu einer Verschlechterung der Schätzgenauigkeit führen (im Vergleich zum nicht-adaptiven Vorgehen). Eine interessante Frage für zukünftige Forschungen ist daher, diese Vorgehensweise zu erweitern oder gar die optimale Aufteilung analytisch zu bestimmen. Weiterhin ist die Betrachtung „optimaler“ Gewichte, welche den mittleren quadratischen Fehler des Model-Averaging Schätzers minimieren, von großem Interesse (siehe etwa Hansen (2007), Liang et al. (2011)). Diese hängen jedoch von den Daten und den speziellen betrachteten Modellen ab und können oftmals nur unter großem Aufwand analytisch bestimmt werden. Ein weiterer Ansatz in diese Richtung ist die Verwendung eines Minimax-Ansatzes für die Wahl der Modellgewichte, um so den mittleren quadratischen Fehler für die „schlechteste“ Wahl von Modellgewichten zu minimieren.

Literaturverzeichnis

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N. und Csáki, F. (Hrsg.), *Proceedings of the Second International Symposium on Information Theory*, 267-281. Akadémiai Kiadó, Budapest, Ungarn.
- Alhorn, K., Schorning, K. und Dette, H. (2019a). Optimal designs for frequentist model averaging. Erscheint in: *Biometrika*.
- Alhorn, K., Dette, H. und Schorning, K. (2019b). Optimal designs for model averaging in non-nested models. SFB 823 Discussion Paper 7/2019.
- Aoki, Y., Röshammar, D., Hamrén, B. und Hooker, A.C. (2017). Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. *Journal of Pharmacokinetics and Pharmacodynamics* **44(6)** 581-597.
- Atkinson, A.C., Donev, A. und Randall, T. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford, Großbritannien.
- Atkinson, A.C. (2008). DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference* **138**, 56-64.
- Atkinson, A.C. und Fedorov, V.V. (1975). The design of experiments for discriminating between two rival models. *Biometrika* **62(1)**, 57-70.
- Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C.-H., König, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro, J.C., Roy, A., Sax, R. und Shen, F. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics* **17(6)**, 965-995.
- Breiman, L. (1996). Bagging predictors. *Machine Learning* **24(2)**, 123-140.
- Bretz, F., Dette, H. und Pinheiro, J.C. (2010). Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine* **29**, 731-742.
- Bretz, F., Pinheiro, J.C. und Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61(3)**, 738-748.

- Bretz, F., Hsu, J. und Pinheiro, J.C. (2008). Dose finding - a challenge in statistics. *Biometrical Journal* **50(4)**, 480-504.
- Buatois, S., Ueckert, S., Frey, N., Retout, S. und Mentré, F. (2018). Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *The AAPS Journal* **20(3)**, 56.
- Buckland, S.T., Burnham, K.P. und Augustin, N.H. (1997). Model selection: an integral part of inference. *Biometrics* **53(2)**, 603-618.
- Burnham, K.P. und Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach*. 2. Auflage, Springer, New York, USA.
- Chaloner, K. und Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science* **10(3)**, 273-304.
- Chen, L., Wan, A.T.K., Tso, G. und Zhang, X. (2018). A model averaging approach for the ordered probit and nested logit models with applications. *Journal of Applied Statistics* **45(16)**, 3012-3052.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics* **24(4)**, 586-602.
- Claeskens, G. und Hjort, N.L. (2003): The focused information criterion. *Journal of the American Statistical Association* **98(464)**, 900-916.
- Claeskens, G. und Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, Großbritannien.
- Cornish-Bowden, A. (2012). *Fundamentals of Enzyme Kinetics*. 4. Auflage, Wiley-VCH, Weinheim, Deutschland.
- Dette, H., Bornkamp, B. und Bretz, F. (2013). On the efficiency of two-stage response-adaptive designs. *Statistics in Medicine* **32**, 1646-1660.
- Dette, H., Bretz, F., Pepelyshev, A. und Pinheiro, J.C. (2008). Optimal designs for dose-finding studies. *Journal of the American Statistical Association* **103(483)**, 1225-1237.
- Dette, H., Kiss, C., Bevanda, M. und Bretz, F. (2010): Optimal designs for the emax, log-linear and exponential models. *Biometrika* **97(2)**, 513-518.
- Dette, H. und Titoff, S. (2009). Optimal discrimination designs. *The Annals of Statistics* **37(4)**, 2056-2082.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York, USA.
- Fedorov, V.V. und Sergej, L. (2013). *Optimal Design for Nonlinear Response Models*. CRC Press, Boca Raton, USA.

- Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M. und Maire, P. (2008). The hill equation: a review of its capabilities in pharmacological modelling. *Fundamental and Clinical Pharmacology* **22**, 633-648.
- Hansen, B.E. (2007). Least squares model averaging. *Econometrica* **75(4)**, 1175-1189.
- Hjort, N.L. und Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association* **98(464)**, 879-899.
- Hoeting, J.A., Madigan, D., Raftery, A.E. und Volinsky, C.T. (1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Statistical Science* **14(4)**, 382-417.
- Holford, N.H. und Sheiner, L.B. (1981). Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. *Clinical Pharmacokinetics* **6**, 429-453.
- Kapetanios, G., Labhard, V. und Price, S. (2008). Forecasting using bayesian and information-theoretic model averaging. *Journal of Business & Economic Statistics* **26(1)**, 33-41.
- Kiefer, J. (1974). General equivalence theory for optimum design (approximate theory). *The Annals of Statistics* **2**, 849-879.
- Kiefer, J. und Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**, 363-366.
- Klenke, A. (2013). *Wahrscheinlichkeitstheorie*. 3. Auflage, Springer, Berlin Heidelberg, Deutschland.
- Läuter, E. (1974). Experimental design in a class of models. *Mathematische Operationsforschung und Statistik* **5(4-5)**, 379-398.
- Leeb, H. und Pötscher, B.M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Annals of Statistics* **34(5)**, 2554-2591.
- Lehmann, E.L. (1983). *Theory of Point Estimation*. John Wiley & Sons, New York, USA.
- Liang, H., Zou, G., Wan, A.T.K. und Zhang, X. (2011). Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association* **106(495)**, 1053-1066.
- López-Fidalgo, J., Tommasi, C. und Trandafir, P.C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society, Series B* **69(2)**, 231-242.
- Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Chichester, Großbritannien.

- MacDougall, J. (2006). Analysis of dose-response studies - emax model. In Ting, N. (Hrsg.), *Dose Finding in Drug Development*, 127-145. Springer, New York, USA.
- Papageorgiou, N.S. und Winkert, P. (2018). *Applied Nonlinear Functional Analysis*, De Gruyter, Berlin, Deutschland.
- Pinheiro, J.C., Bornkamp, B. und Bretz, F. (2006). Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics* **16(5)**, 639-656.
- Pinheiro, J.C., Bornkamp, B., Glimm, E. und Bretz, F. (2014). Model-based dose finding under model uncertainty using general parametric models. *Statistics in Medicine* **33(10)**, 1646-1661.
- Pötscher, B.M. (1991). Effects of model selection on inference. *Econometric Theory* **7**, 163-185.
- Powell, M.J.D. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In Hennart, J.-P. and Gomez, S. (Hrsg.), *Advances in Optimization and Numerical Analysis*, 51-67. Kluwer Academic Publishers, Dordrecht, Niederlande.
- Pshenichnyi, B.N. (1971). *Necessary Conditions for an Extremum*. Marcel Dekker, Inc., New York, USA.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York, USA.
- Pukelsheim, F. und Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79(4)**, 763-770.
- Raftery, A.E. und Zheng, Y. (2003). Discussion: performance of bayesian model averaging. *Journal of the American Statistical Association* **98(464)**, 931-938.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. John Wiley & Sons, New York, USA.
- Schorning, K., Bornkamp, B., Bretz, F. und Dette, H. (2016). Model selection versus model averaging in dose finding studies. *Statistics in Medicine* **35**, 4012-4040.
- Schumacher, M. und Schulgen, G. (2007). *Methodik klinischer Studien: Methodische Grundlagen der Planung, Durchführung und Auswertung*. 2. Auflage, Springer, Berlin, Deutschland.
- Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall, London, Großbritannien.
- Stock, J.H. und Watson, M.W. (2004). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, **41(3)**, 788-829.
- Tommasi, C. (2009). Optimal designs for both model discrimination and parameter estimation. *Journal of Statistical Planning and Inference* **139(12)**, 4123-4132.

- Uciński, D. und Bogacka, B. (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society Series B* **67(1)**, 3-18.
- Van der Vaart, A.W. (2007). *Asymptotic Statistics*. 8. Auflage, Cambridge University Press, Cambridge, Großbritannien.
- Verrier, D., Sivapregassam, S. und Solente, A.-C. (2014). Dose-finding studies, mcp-mod, model selection, and model averaging: Two applications in the real world. *Clinical Trials*, **11**, 476-484.
- Wagner, M. and Hlouskova, J. (2015). Growth regressions, principal components augmented regressions and frequentist model averaging. *Jahrbücher für Nationalökonomie und Statistik* **235(6)**, 642-662.
- Werner, D. (2018). *Funktionalanalysis*, 8. Auflage, Springer, Berlin, Heidelberg, Deutschland.
- White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* **76(374)**, 419-433.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50(1)**, 1-25.
- Wiens, D. (2015). Robustness of design. *Journal of the Royal Statistical Society, Series B* **71(4)**, 805-829.
- Woolridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2. Auflage, MIT Press, Cambridge, Großbritannien.
- Zen, M.-M. und Tsai, M.-H. (2002). Some criterion-robust optimal designs for the dual problem of model discrimination and parameter estimation. *Sankhya: The Indian Journal of Statistics* **64**, 322-338.
- Zwietering, M.H., Jongenburger, I., Rombouts, F.M. und Van 't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology* **56(6)**, 1875-1881.