

Bestimmung c-optimaler Versuchspläne in Modellen mit
zufälligen Effekten, mit Anwendungen in der
Pharmakokinetik

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund.

Der Fakultät Statistik vorgelegt von

Tim Holland-Letz

aus Bochum

15. Dezember 2009

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden.

Bochum, der 15.12.2009,

Tim Holland-Letz

Gutachter:

Prof. Dr. J.Kunert

Prof. Dr. H.Dette

Prof. Dr. C. Müller

Tag der mündlichen Prüfung: 8.März 2010

Summary

Clinical experiments in the area of pharmacokinetics are often based on random effects models (so called population models with repeated measurements on several patients each). This situation poses methodological challenges for optimal experimental design, as the setup results both in partly correlated observations and a parameter dependant variance, thus violating two key assumptions of classical design theory. It is the aim of this thesis to adapt and extend the existing design methods to cover some of these situations. As the most important pharmacokinetic parameters are summary statistics of several model parameters (e.g. the area under the concentration curve) the focus will be on c -optimal designs which allow optimal estimations of these quantities. First, the geometric representation proposed by Elfving (1952) is generalized to allow for the situation described above. Second, the equivalence theory by Kiefer (1974) and Pukelsheim (1993) is applied to the specific situation. Third, multiplicative algorithms for numerical computation of optimal designs are adapted, and finally the concept of asymptotically optimal designs is applied to a specific case of correlated observations.

Überblick

Medizinische Studien im Bereich der Pharmakokinetik basieren in vielen Fällen auf speziellen Modellen mit zufälligen Effekten, den sogenannten Populationsmodellen. In solchen Studien werden jeweils mehrere Messungen an einer Anzahl verschiedener Patienten durchgeführt. Aus dem Blickwinkel der optimalen Versuchsplanung führt dies zu methodischen Schwierigkeiten, da der zufällige Effekt sowohl eine parameterabhängige Varianz der Beobachtungen als auch teilweise korrelierte Daten zur Folge hat. Auf diese Weise sind zwei der Schlüsselannahmen klassischer Versuchsplanungsliteratur verletzt. Es ist das Ziel dieser Arbeit, die bestehende Methodik so anzupassen und zu ergänzen, dass auch diese Situationen betrachtet werden können. Da die wichtigsten zu schätzenden Kenngrößen in der Pharmakokinetik bestimmte Summengrößen der Parameter sind (z.B. die Fläche unter der Konzentrationskurve eines Präparates) wird der Schwerpunkt dieser Arbeit auf c -optimalen Designs liegen, die die optimale Schätzung solcher Größen erlauben.

Im einzelnen wird zunächst die geometrische Repräsentation optimaler Designs nach Elfving (1952) so verallgemeinert, dass die beschriebene Situation abgedeckt ist. Im zweiten Schritt wird die Äquivalenztheorie nach Kiefer (1974) und Pukelsheim (1993) auf das spezifische Modell angewendet. Dritter Schritt ist die Anpassung multiplikativer Algorithmen zur numerischen Bestimmung optimaler Designs in dieser Situation, und als letztes Ergebnis wird das Konzept asymptotisch optimaler Designs auf einen speziellen Fall korrelierter Beobachtungen angewendet.

Inhaltsverzeichnis

1	Einleitung	7
2	Modellbildungen	10
2.1	Modelle mit fixen Effekten	10
2.1.1	Lineare Modelle	10
2.1.2	Nichtlineare Modelle	11
2.2	Modelle mit zufälligen Effekten	11
2.3	Populationsmodelle	13
2.4	Modelle der Pharmakokinetik	14
3	Grundlagen der optimalen Versuchsplanung	17
3.1	Schätzverfahren	17
3.1.1	Kleinste-Quadrate-Schätzer	17
3.1.2	Maximum-Likelihood-Schätzer	19
3.2	Fisher Informationen für konkrete Fehlerverteilungen	20
3.2.1	Fisher-Information bei normalverteilten Störgrößen	21
3.2.2	Fisher-Information bei lognormalverteilten Fehlern	24
3.3	Exakte und approximative Designs	25
3.4	Informationsmatrizen	26
3.4.1	Kleinste-Quadrate-Schätzer	27

3.4.2	Maximum Likelihood Schätzer	30
3.5	Informationsfunktionen	32
3.5.1	Schätzbarkeit	33
3.5.2	ϕ^p -Optimalität	34
3.5.3	c-Optimalität für abgeleitete Kennzahlen der Pharmakokinetik	35
4	Äquivalenzsätze	38
4.1	Der allgemeine Äquivalenzsatz für ϕ^p -Optimalität	38
4.2	Beispiele	40
5	Elfving-Sätze	43
5.1	Satz von Elfving	43
5.1.1	Beispiel im Modell mit fixen Effekten	44
5.2	Verallgemeinerter Satz von Elfving	45
5.3	Beispiele	50
5.3.1	ML-Schätzer im Modell mit zufälligen Effekten	50
5.3.2	NWLS-Schätzer im linearen Modell mit korrelierten Beobachtungen	54
5.3.3	NWLS-Schätzer im Populationsmodell	59
6	Algorithmen	62
6.1	Ein multiplikativer Algorithmus für c-optimale Designs	62
6.1.1	Beispiel: Unkorrelierte Beobachtungen	65
6.1.2	Beispiel: Korrelierte Beobachtungen	66
6.1.3	Beispiel: ML-Schätzer bei korrelierten Beobachtungen	70
7	Asymptotisch optimale Designs bei korrelierten Beobachtungen	72
7.1	Der Nelder-Mead Algorithmus	73
7.2	Asymptotisch effiziente Designs	74

7.2.1 Beispiele	79
---------------------------	----

8 Zusammenfassung	84
--------------------------	-----------

Kapitel 1

Einleitung

Das Prinzip der optimalen Versuchsplanung befasst sich mit dem Problem, Rahmenbedingungen eines wissenschaftlichen Experimentes so festzulegen, dass das Ziel des Experimentes bestmöglich erreicht werden kann. Üblicherweise ist der Zweck solcher Experimente, bestimmte Parameter realer Zusammenhänge zu schätzen. Sind diese Parameter Kenngrößen von stetigen Funktionen der Rahmenbedingungen, so spricht man von stetiger optimaler Versuchsplanung, und diese ist der Gegenstand dieser Arbeit.

Als praktischer Hintergrund dienen Problemstellungen aus der Medizin, speziell der klinischen Pharmakokinetik. Die beschriebenen Methoden sind jedoch in der Regel problemlos auf Anwendungen aus anderen Bereichen übertragbar. Die Bestimmung der Pharmakokinetik medizinischer Präparate ist ein Schritt in der Entwicklung medizinischer Wirkstoffe, üblicherweise in Phase 1 und 2 der klinischen Prüfung, nach Tierversuchen aber vor der Durchführung umfangreicher Wirksamkeitsstudien an echten Patienten. Ziel ist es, den Verlauf der Konzentration des Präparates in relevanten Bereichen des Körpers zu modellieren. Üblicherweise geschieht dies über die Theorie der Kompartimentmodelle (Shargel (1993) oder Källén (2008) für technische Aspekte), und die Wirkstoffkonzentration zu einem bestimmten Zeitpunkt x wird als Funktion der Zeit und einer Anzahl von Parametern beschrieben. Der Typus der zu verwendenden Funktion ist aufgrund der Modellvorstellung bis auf Parameter durch die Art der Verabreichung und das Wirkprinzip vorgegeben. Die Wirkstoffe werden dann in klinischen Studien an einer Anzahl verschiedener

Patienten geprüft. Im Gegensatz zu klassischen Ansätzen, die versuchen, für jeden einzelnen Patienten den kompletten Parametersatz zu schätzen, haben modernere Studien üblicherweise das Ziel, globale Erwartungswerte der Parameter über die Population der möglichen Patienten zu schätzen. Das resultierende Modell ist ein Modell mit zufälligen Effekten, das Populationsmodell (siehe z.B. Sheiner und Wakefield (1999)). Im Gegensatz zur individuellen Parameterschätzung reichen hier wenige Messungen pro individuellem Patienten aus.

Das primäre Interesse der Studie ist häufig nicht die eigentliche Parameterschätzung, sondern die Schätzung von abgeleiteten Summengrößen der Pharmakokinetik wie z.B. die Gesamtfläche unter der Konzentrationskurve (die AUC) als Maß für den Gesamteffekt eines Präparates, oder der Zeitpunkt und die Ausprägung der maximalen Wirkstoffkonzentration. Solche Schätzprobleme entsprechen im Kontext der Versuchplanung den c -optimalen Designproblemen und diese sind daher der Hauptfokus dieser Arbeit. Grundidee ist es, einen Versuchsplan für die Wahl der Zeitpunkte der verfügbaren Konzentrationsmessungen an Patienten so festzulegen, dass die gesuchten Summengrößen mit minimaler Varianz geschätzt werden können.

Zum jetzigen Zeitpunkt existieren bereits umfangreiche Erkenntnisse zum Prinzip der c -optimalen Versuchsplanung (siehe z.B. Pukelsheim (1993), Pázman (1986), Elfving (1952) für die allgemeine Theorie oder Karlin und Studden (1966) für spezifische Situationen). Die speziellen Problemstellungen der Pharmakokinetik weisen jedoch einige Eigenheiten auf, die in der klassischen Versuchsplanungsliteratur nur unzureichend betrachtet werden. So führt die Verwendung von Populationsmodellen zum einen zu korrelierten Messungen bei Beobachtungen am selben Patienten und zum anderen dazu, dass auch die Varianz der Beobachtungen abhängig von den zu schätzenden Parametern ist. Dies hat Auswirkungen auf bestimmte Schätzverfahren (speziell Maximum-Likelihood-Schätzungen) und sollte daher in der Versuchsplanung berücksichtigt werden. Ziel dieser Arbeit ist es, zunächst soweit möglich existierende Erkenntnisse der Versuchsplanung auf Probleme der Pharmakokinetik anzuwenden und anschließend Lösungsansätze für die beschriebenen Problematiken aufzuzeigen.

In Kapitel 2 werden zunächst die verwendeten Modelle vorgestellt, während in Kapitel 3 die gängigen Schätzverfahren ebenso wie Kriterien zur Bewertung eines Versuchsplans eingeführt werden. Kapitel 4 beinhaltet Resultate zur Bestätigung der Optimalität eines gegebenen Ver-

suchsplans. Kapitel 5 liefert einige graphische Ansätze zur Bestimmung und Visualisierung solcher Versuchspläne, während Kapitel 6 und 7 numerische Prozeduren zum Erreichen dieses Ziels behandeln.

Kapitel 2

Modellbildungen

2.1 Modelle mit fixen Effekten

2.1.1 Lineare Modelle

Das einfachste in dieser Arbeit betrachtete Modell ist ein lineares Regressionsmodell der Form

$$Y_i = f(x_i)^T \theta + \epsilon_i; \quad i = 1, \dots, n, \quad (2.1.1)$$

Der Vektor $Y = (Y_1, \dots, Y_n)^T$ stellt n verschiedene Beobachtungen eines Experimentes dar, der Vektor der Designvariablen $x = (x_1, \dots, x_n)^T$ bezeichne die festzulegenden Versuchsbedingungen zu jeder Beobachtung x_i , wählbar aus einem kompakten Versuchsbereich $\mathcal{X} \subset \mathbb{R}^k$, $i = 1, \dots, n$. Der Vektor $\theta = (\theta_1, \dots, \theta_k)^T \in \Theta \subset \mathbb{R}^k$ bezeichne zu schätzende Parameter und $f(x) : \mathcal{X} \rightarrow \mathbb{R}^k$ den Spaltenvektor einer k -dimensional vektorwertigen bekannten zweimal stetig differenzierbaren Funktion mit linear unabhängigen Komponenten. Die $k \times n$ -Matrix

$$F := (f(x_1), \dots, f(x_n))$$

sei als Designmatrix bezeichnet. Für die zufälligen Störgrößen $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ gelte $\epsilon \sim \mathcal{N}(0, W)$, mit $W \in \mathbb{R}^{n \times n}$ positiv semidefinit. Die Kovarianzmatrix W sei zunächst als Diagonalmatrix $diag(\sigma_1^2, \dots, \sigma_n^2)$ angenommen, $\sigma_i > 0, i = 1 \dots n$.

2.1.2 Nichtlineare Modelle

Eine Verallgemeinerung des linearen Modells erlaubt auch nichtlineare Beziehungen, d.h. ein Modell der Form

$$Y_i = \eta(x_i, \theta) + \epsilon_i \quad i = 1, \dots, n, \quad (2.1.2)$$

wobei $\eta(x_i, \theta) : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ eine beliebige, aber bekannte und zweimal stetig differenzierbare Funktion bezeichne.

Zum Zwecke der Versuchsplanung ist es sinnvoll, dieses Modell durch eine lineare Taylor-Approximation erster Ordnung anzunähern:

$$\eta(x, \theta) \approx \eta(x, \theta_0) + \tilde{f}(x, \theta_0)^T (\theta - \theta_0), \quad (2.1.3)$$

mit dem Spaltenvektor

$$\tilde{f}(x, \theta) = \frac{\partial \eta(x, \theta)}{\partial \theta}$$

d.h.

$$Y_i \approx \eta(x_i, \theta_0) + \tilde{f}(x, \theta_0)^T (\theta - \theta_0) + \epsilon_i \quad i = 1, \dots, n, \quad (2.1.4)$$

Die Funktionen \tilde{f} sind in diesem Fall abhängig vom Parameter θ_0 . Die Varianz der Beobachtungen Y_i entspricht der Varianz der Störgrößen, d.h. $Var(Y) = W$.

2.2 Modelle mit zufälligen Effekten

Modelle mit zufälligen Effekten sind Modelle, in denen Beobachtungen nicht von konstanten, unveränderlichen Parametern abhängen, sondern diese Parameter selbst für jede Beobachtung Realisationen von Zufallsvariablen mit bestimmter Verteilung sind. Dabei sei für diesen Abschnitt angenommen, dass verschiedene Beobachtungen stochastisch unabhängig sind. Modelle mit gleicher Realisation der Zufallseffekte bei mehr als einer Beobachtung werden im Rahmen

der Populationsmodelle in Abschnitt 2.3 betrachtet.

Die Modelle sind dann die folgenden:

$$Y_i = f(x_i)^T b_i + \epsilon_i; \quad i = 1, \dots, n, \quad (\text{linear}) \quad (2.2.1)$$

$$Y_i = \eta(b_i, x_i) + \epsilon_i; \quad i = 1, \dots, n, \quad (\text{allgemein}) \quad (2.2.2)$$

Auch hier sind Y_i die n verschiedenen Beobachtungen, die Designvariablen $x_i \in \mathcal{X} \subset \mathbb{R}$, ($i = 1, \dots, n$) können durch den Versuchsplaner festgelegt werden und $\epsilon \sim \mathcal{N}(0, W)$. Der Vektor $b_i = (b_{i1}, \dots, b_{ik})^T$ der individuellen Parameter sei für jede Beobachtung Resultat eines Zufallsexperimentes mit Verteilung $\mathcal{N}(\theta, \Omega)$, unabhängig von den Störgrößen ϵ_i . Ziel des Versuches ist es, den globalen Erwartungswertvektor $\theta \in \Theta \subset \mathbb{R}^k$ und in manchen Fällen die Kovarianzmatrix $\Omega \in \mathbb{R}^{k \times k}$ zu schätzen.

Wie zuvor ist für das allgemeine Modell eine Taylor-Annäherung hilfreich:

$$Y_i \approx \eta(x_i, \theta) + \tilde{f}(x_i, \theta)^T (b_i - \theta) + \epsilon_i \quad i = 1, \dots, n, \quad (2.2.3)$$

mit

$$\tilde{f}(x_i, \theta) = \frac{\partial \eta(x_i, \theta)}{\partial \theta}$$

Die Varianz der Zufallsvariablen Y_i , $i = 1, \dots, n$, kann damit angenähert werden durch

$$\text{Var}(Y_i) \approx \tilde{f}(x_i, \theta)^T \Omega \tilde{f}(x_i, \theta) + \sigma_i^2. \quad (2.2.4)$$

Bezeichne jetzt analog wie zuvor

$$\tilde{F} := (\tilde{f}(x_1, \theta), \dots, \tilde{f}(x_n, \theta)) \quad (2.2.5)$$

die Designmatrix, dann kann die Varianz des gesamten Zufallsvektors $Y = (Y_1, \dots, Y_n)^T$ dargestellt werden als

$$V := \text{Var}(Y) \approx \text{diag}(\tilde{F}^T \Omega \tilde{F}) + W. \quad (2.2.6)$$

Dies gilt auch dann, wenn W keine Diagonalmatrix ist.

Im Spezialfall des linearen Modells (2.2.1) gilt:

$$V = \text{Var}(Y) = \text{diag}(F^T \Omega F) + W. \quad (2.2.7)$$

2.3 Populationsmodelle

In praktischen medizinischen Fragestellungen der Pharmakokinetik ist es in der Regel das Ziel, Parameter bestimmter Konzentrationsfunktionen anhand von Studien mit wiederholten Messungen an verschiedenen Probanden zu schätzen. Dabei ist es üblicherweise nicht von Interesse, die individuellen Parameter eines konkreten einzelnen Probanden zu bestimmen, und es kann weiterhin nicht davon ausgegangen werden, dass diese Parameter für alle Probanden identisch sind. Daher wird in dieser Situation ein Modell mit zufälligen Effekten angenommen, mit der Absicht, die globalen Erwartungswerte der Parameter zu schätzen. Dieses Modell wird als Populationsmodell bezeichnet [Retout und Mentré (2003)].

Im Unterschied zum Modell in 2.2 stehen dabei $m > 1$ (möglicherweise korrelierte) Messungen pro Proband zu Verfügung, ebenso wie insgesamt N Probanden. Die Gesamtzahl an Messungen wird wie zuvor mit $n = Nm$ bezeichnet, und man erhält die folgende Darstellung:

$$Y_{ij} = \eta(x_{ij}, b_i) + \varepsilon_{ij} \quad i = 1, \dots, N, j = 1, \dots, m, \quad (2.3.1)$$

Dazu sei $\eta : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ eine bekannte, zweimal stetig differenzierbare und nicht notwendig lineare Funktion, die Fehler $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{im})^T$ für jeden Patienten normalverteilt mit Erwartungswert 0 und Kovarianzmatrix $W_i \in \mathbb{R}^{m \times m}$, $i = 1, \dots, N$. Die Zufallsvariablen $b_1, \dots, b_N \sim \mathcal{N}(\theta, \Omega)$ seien k -dimensional unabhängig normalverteilt mit Erwartungswert $\theta \in \Theta \subset \mathbb{R}^k$ und Kovarianzmatrix $\Omega \in \mathbb{R}^{k \times k}$ und repräsentieren den Effekt des zugehörigen Individuums [siehe Beatty und Piegorsch (1997), Ette et al. (1995), Cayen und Black (1993)]. Weiterhin wird stochastische Unabhängigkeit der Zufallsvariablen b_1, \dots, b_N und $(\varepsilon_{11}, \dots, \varepsilon_{Nm})^T$ angenommen.

Auch hier ist zum Zwecke der Versuchsplanung eine Linearisierung des Modells sinnvoll:

$$Y_{ij} \approx \eta(x_{ij}, \theta) + \tilde{f}(x_{ij}, \theta)^T (b_i - \theta) + \varepsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, m, \quad (2.3.2)$$

mit

$$\tilde{f}(x, b) = \frac{\partial \eta(x, b)}{\partial b}.$$

Der gemeinsame zufällige Effekt aller Beobachtungen eines Probanden führt dazu, dass die Beobachtungen an diesem Individuum korreliert sind. Fasst man alle Beobachtungen an einem

Individuum zu einem Vektor $Y_i = (Y_{i1}, \dots, Y_{im})^T$ zusammen und definiert analog eine individuelle Designmatrix durch

$$\tilde{F}_i := (\tilde{f}(x_{i1}, \theta), \dots, \tilde{f}(x_{im}, \theta)) \quad i = 1, \dots, N, \quad (2.3.3)$$

so kann die Kovarianzmatrix wie folgt dargestellt werden:

$$V_i := \text{Var}(Y_i) \approx \tilde{F}_i^T \Omega \tilde{F}_i + W_i \quad i = 1, \dots, N. \quad (2.3.4)$$

Das Modell kann dann als gewöhnliches Modell mit zufälligen Effekten (2.2.2) betrachtet werden, mit $Y = (Y_1^T, \dots, Y_N^T)^T$ und

$$V = \text{Var}(Y) = \text{diag}(V_1, \dots, V_N). \quad (2.3.5)$$

2.4 Modelle der Pharmakokinetik

Praktischer Hintergrund der hier beschriebenen Überlegungen sind Fragestellungen der medizinischen Pharmakokinetik, speziell Modellierungen der Beziehung zwischen Zeit und Wirkstoffkonzentration im Körper nach Gabe eines Medikamentes. Grundlegende Vorstellung ist dabei, dass der menschliche Körper aus einem oder mehreren homogenen Bereichen (Kompartments) mit zu einem fixen Zeitpunkt gleicher Wirkstoffkonzentration besteht. Über die Zeit findet ein Aufbau (Invasion) und Abbau (Evasion) der Konzentration für jedes Kompartment statt. Je nach Anzahl dieser homogenen Bereiche werden diese Ansätze als Ein- oder Mehrkompartimentmodelle bezeichnet (Shargel (1993), Kapitel 4-5, Rowland (1993), Kapitel 19, Beatty und Piegorsch (1997)).

Wichtige Parameter zur Beschreibung solcher Konzentrationsverläufe sind die folgenden:

D die Dosis des verabreichten Präparates.

V das Volumen des entsprechenden Körperbereiches (Kompartments)

k_c die Abbaukonstante des Präparates (Anteil pro Zeiteinheit)

k_a die Aufnahmekonstante des Präparates

$C(x)$ die Konzentration des Präparates zum Zeitpunkt x

Im einfachsten Fall der intravenösen Injektion des Präparates im Ein-Kompartiment-Modell erreicht die Konzentration $C(x)$ des Präparates nahezu sofort nach Verabreichung ihr Maximum und fällt dann im Zeitverlauf exponentiell ab. Die Konzentrationsfunktion ist dann die folgende (vgl. Retout et al. (2001)):

$$C(x) = \frac{D}{V} e^{-k_c x} \quad (2.4.1)$$

Man nennt dieses Modell auch das Modell der einfachen Evasion.

Im Falle des Ein-Kompartiment-Modells mit oraler Verabreichung entsteht die Konzentrationsfunktion anhand zweier paralleler Prozesse, der Aufnahme des Präparates durch den Verdauungstrakt und dem simultanen Abbau der bereits vorhandenen Stoffmenge im Körper. Die Konzentrationsfunktion ergibt sich daher als Lösung der Differentialgleichung dieser beiden simultanen Prozesse. Sie wird als Bateman-Funktion bezeichnet und hängt von 4 Parametern ab (Atkinson et al. (1993), Jones et al. (1999)):

$$C(x) = \frac{Dk_a}{V(k_a - k_c)} (e^{-k_c x} - e^{-k_a x}) \quad (2.4.2)$$

Der Parameter D (Dosis) ist in beiden Modellen bekannt und muss daher nicht geschätzt werden.

Bei Mehr-Kompartiment-Modellen können die Konzentrationsfunktionen in einem bestimmten Kompartiment allgemein als Summe von Exponentialfunktionen der folgenden Art dargestellt werden ($\eta : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^+$, $\theta = (\theta_1, \dots, \theta_{2r}) \in \Theta \subset \mathbb{R}^{2r}$):

$$\eta(x, \theta) = \sum_{i=1}^r \theta_i e^{-\theta_{i+r} x} \quad (2.4.3)$$

Im Regelfall der intravenösen Injektion des Präparates ist für jedes Kompartiment ein Summen-term notwendig. Der Spezialfall $r = 1$ beschreibt demnach bei entsprechender Parameterwahl das Modell der einfachen Evasion

$$\eta(x, \theta) = \theta_1 e^{-\theta_2 x}, \quad (2.4.4)$$

und der Spezialfall $r = 2$ und $\theta_2 = -\theta_1$ die orale Verabreichung im Ein-Kompartiment-Modell, reparametrisiert darstellbar als:

$$\eta(x, \theta) = \theta_1(e^{-\theta_2 x} - e^{-\theta_3 x}) \quad (2.4.5)$$

Im Weiteren werden statt der Darstellungen (2.4.1) und (2.4.2) nur die allgemeinen Darstellungen (2.4.4) und (2.4.5) verwendet. Eine graphische Darstellung dieser Konzentrationsfunktionen ist in Abbildung 2.1 zu finden.

Zwei- und Mehrkompartimentmodelle sind denkbar, in der Praxis aber von geringer Bedeutung.

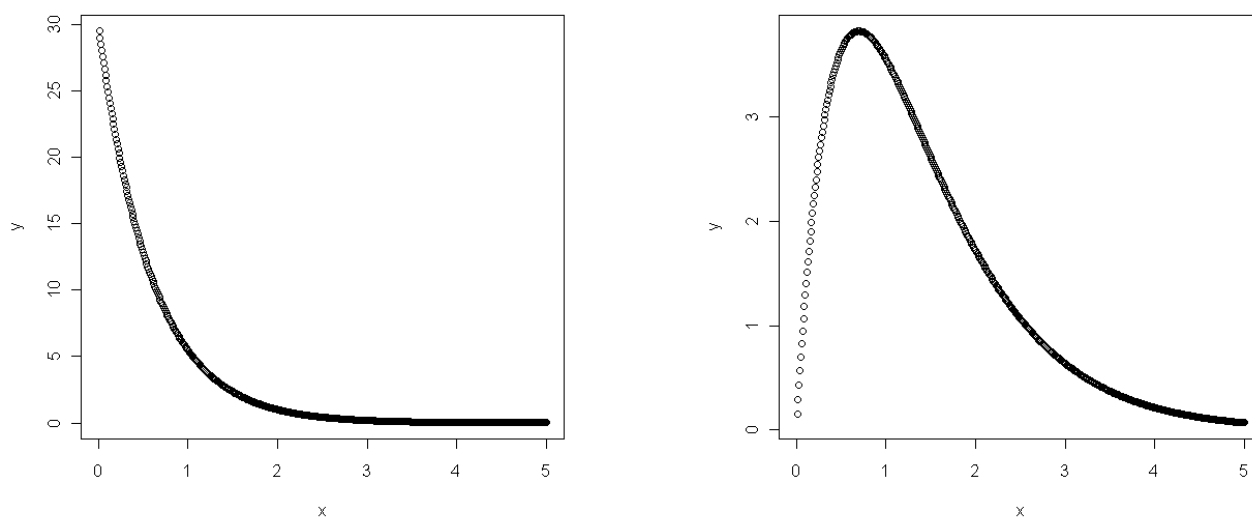


Abbildung 2.1: Graphische Darstellung der Evasions- (2.4.4) und der Bateman-Funktion (2.4.5).

Kapitel 3

Grundlagen der optimalen Versuchsplanung

3.1 Schätzverfahren

Ziel der optimalen Versuchsplanung ist es, in den beschriebenen Modellen die Rahmenbedingungen von Experimenten so festzulegen, dass die Parameter θ mit minimaler Varianz geschätzt werden können. Zu diesem Zweck ist es notwendig, zunächst die gängig verwendeten Schätzverfahren (Kleinste-Quadrate- und Maximum-Likelihood-Schätzer) einzuführen.

3.1.1 Kleinste-Quadrate-Schätzer

Wir betrachten zunächst den linearen Fall.

Definition 3.1

Gegeben sei ein Vektor von Realisationen $y = (y_1, \dots, y_n)^T$ aus Modell (2.2.1) mit bekannter Kovarianzmatrix $V = \text{diag}(F^T \Omega F) + W$ wie in (2.2.7) definiert und Versuchsbedingungen $x = (x_1, \dots, x_n)^T$. Ein Schätzer $\hat{\theta}$ für θ heißt gewöhnlicher Kleinste-Quadrate-Schätzer (Ordinary-

Least-Squares, OLS), falls für alle $\theta \in \Theta$ gilt, dass

$$(y - F^T \hat{\theta})^T (y - F^T \hat{\theta}) \leq (y - F^T \theta)^T (y - F^T \theta). \quad (3.1.1)$$

Ein Schätzer $\hat{\theta}$ für θ heißt gewichteter Kleinste-Quadrate-Schätzer (Weighted-Least-Squares, WLS), falls für alle $\theta \in \Theta$ gilt, dass

$$(y - F^T \hat{\theta})^T V^{-1} (y - F^T \hat{\theta}) \leq (y - F^T \theta)^T V^{-1} (y - F^T \theta). \quad (3.1.2)$$

Der gewöhnliche Kleinste-Quadrate-Schätzer in Modell 2.2.1 ist gegeben durch (vgl. Searle (1971), Kap. 3.3)

$$\hat{\theta}_{OLS} = (FF^T)^{-1} Fy, \quad (3.1.3)$$

mit Varianz

$$\text{Var}(\hat{\theta}_{OLS}) = \text{Var}((FF^T)^{-1} Fy) = (FF^T)^{-1} FV F^T (FF^T)^{-1}. \quad (3.1.4)$$

Der gewichtete Kleinste-Quadrate-Schätzer im selben Modell ist gegeben durch

$$\hat{\theta}_{WLS} = (FV^{-1} F^T)^{-1} FV^{-1} y. \quad (3.1.5)$$

Die Varianz des WLS-Schätzers ist damit gegeben durch

$$\text{Var}(\hat{\theta}_{WLS}) = \text{Var}((FV^{-1} F^T)^{-1} FV^{-1} y) = (FV^{-1} F^T)^{-1} \quad (3.1.6)$$

Falls $V = \sigma^2 I$ für beliebige Konstante $\sigma > 0$ gilt $\text{Var}(\hat{\theta}_{WLS}) = \text{Var}(\hat{\theta}_{OLS}) = (FV^{-1} F^T)^{-1} = \sigma^2 (FF^T)^{-1}$.

Auf analoge Weise ist es auch möglich, Kleinste-Quadrate-Schätzer für Modelle mit nichtlinearen Zusammenhängen zu definieren:

Definition 3.2

Gegeben sei ein Vektor von Realisationen $y = (y_1, \dots, y_n)^T$ aus Modell (2.2.2) mit bekannter Kovarianzmatrix $V \approx \text{diag}(\tilde{F}^T \Omega \tilde{F}) + W$ wie in (2.2.6) definiert und Versuchsbedingungen

$x = (x_1, \dots, x_n)^T$. Bezeichne weiterhin $H(x, \theta) = (\eta(x_1, \theta), \dots, \eta(x_n, \theta))^T$. Ein Schätzer $\hat{\theta}$ für θ heißt nichtlinearer gewöhnlicher Kleinste-Quadrate-Schätzer (Nonlinear-Ordinary-Least-Squares, NOLS), falls für alle $\theta \in \Theta$ gilt, dass

$$(y - H(x, \hat{\theta}))^T (y - H(x, \hat{\theta})) \leq (y - H(x, \theta))^T (y - H(x, \theta)). \quad (3.1.7)$$

Ein Schätzer $\hat{\theta}$ für θ heißt nichtlinearer gewichteter Kleinste-Quadrate-Schätzer (Nonlinear-Weighted-Least-Squares, NWLS), falls für alle $\theta \in \Theta$ gilt, dass

$$(y - H(x, \hat{\theta}))^T V^{-1} (y - H(x, \hat{\theta})) \leq (y - H(x, \theta))^T V^{-1} (y - H(x, \theta)). \quad (3.1.8)$$

Die NOLS- und NWLS-Schätzer können nicht in geschlossener Form dargestellt werden und müssen numerisch ermittelt werden. Auch die Varianz der Schätzer kann nicht analytisch bestimmt werden. Zum Zwecke der Versuchsplanung verwenden wir daher für die Varianz die Linearisierung (2.2.3) und die Varianzen der linearen Least-Squares-Schätzer als Näherungen, d.h.:

$$\text{Var}(\hat{\theta}_{NOLS}) \approx (\tilde{F}\tilde{F}^T)^{-1}\tilde{F}V\tilde{F}^T(\tilde{F}\tilde{F}^T)^{-1}, \quad (3.1.9)$$

und

$$\text{Var}(\hat{\theta}_{NWLS}) \approx (\tilde{F}V^{-1}\tilde{F}^T)^{-1}, \quad (3.1.10)$$

mit V wie in 2.2.6 und \tilde{F} wie in 2.2.5 definiert. Für Simulationsstudien zur Qualität dieser Approximationen, siehe Dette et al. (2008a).

3.1.2 Maximum-Likelihood-Schätzer

Definition 3.3

Gegeben sei ein Vektor von Realisationen $y = (y_1, \dots, y_n)^T$ aus Modell (2.2.2) unter Versuchsbedingungen $x = (x_1, \dots, x_n)^T$. Sei weiterhin $l(y, x, \theta) = P(Y = y|\theta)$ die zugehörige Likelihood-Funktion.

Ein Schätzer $\hat{\theta}$ für θ heißt *Maximum-Likelihood-Schätzer (ML-Schätzer)*, falls für alle $\theta \in \Theta$ gilt, dass

$$l(y, x, \hat{\theta}) \geq l(y, x, \theta). \quad (3.1.11)$$

Die Varianz des ML-Schätzers kann nicht allgemein ermittelt werden. Unter bestimmten Regularitätsbedingungen (Lehmann und Casella (1998), Kapitel 6, Theorem 5.1) kann jedoch gezeigt werden, dass sich die Varianz asymptotisch der Inverse der Fisher-Information $I(y, x, \theta)$ annähert. Diese Darstellung ist abhängig von der zugrundeliegenden Fehlerverteilung und wird in Abschnitt 3.2 behandelt. Zur Diskussion der Regularitätsbedingungen ebenso wie für deren Nachweis für die multivariate Normalverteilung sei auf Lehmann und Casella (1998) oder Cox und Hinkley (2000) verwiesen.

3.2 Fisher Informationen für konkrete Fehlerverteilungen

In Abschnitt 3.1 zeigte sich, dass die Varianz des Maximum-Likelihood-Schätzer annähernd durch die inverse Fisher-Informationsmatrix bestimmt werden kann. Diese wahrscheinlichkeitstheoretische Kenngröße ist durch die folgende Definition gegeben.

Definition 3.4

Gegeben sei ein Modell vom Typ (2.2.2) mit Realisationen $y = (y_1, \dots, y_n)^T$ unter Versuchsbedingungen $x = (x_1, \dots, x_n)^T$ und sei $l(y, x, \tilde{\theta}) = P(Y = y | \theta = \tilde{\theta})$ die zugehörige Likelihood-Funktion. Die Darstellung

$$I_{i,j}(y, x, \theta) = E\left(\frac{\partial \ln l(y, x, \theta)}{\partial \theta_i} \cdot \frac{\partial \ln l(y, x, \theta)}{\partial \theta_j}\right), \quad i, j = 1, \dots, k \quad (3.2.1)$$

bezeichne das i, j -te Element der Fisher-Informationsmatrix $I(y, x, \theta)$ bezüglich der Parameter $\theta = (\theta_1, \dots, \theta_k)^T$.

Unter der Regularitätsbedingung

$$\int_{\mathbb{R}^n} \frac{\partial^2}{\partial \theta^2} l(y, x, \theta) dy = 0 \quad (3.2.2)$$

kann die Fisher-Information alternativ auch berechnet werden durch

$$I_{i,j}(y, x, \theta) = -E\left(\frac{\partial^2 \ln l(y, x, \theta)}{\partial \theta_i \partial \theta_j}\right), \quad i, j = 1, \dots, k \quad (3.2.3)$$

Die genaue Form der Fisher Informationsmatrix hängt von der zugrundeliegenden Verteilung der Messfehler ab.

3.2.1 Fisher-Information bei normalverteilten Störgrößen

Theorem 3.5

Sei $Y = (Y_1, \dots, Y_n)^T$ normalverteilt mit Erwartungswert $\mu(x, \theta) = (\mu_1, \dots, \mu_n)^T = \mu_\theta$ und Kovarianzmatrix $\Sigma(x, \theta) = \Sigma_\theta$, abhängig von Versuchsbedingungen $x = (x_1, \dots, x_n)^T \in \mathcal{X}^n$ und Parametern $\theta = (\theta_1, \dots, \theta_k)^T \in \Theta$. Seien außerdem die Funktionen $\mu(x, \theta)$ und $\Sigma(x, \theta)$ komponentenweise stetig differenzierbar nach θ . Dann haben die Einträge der Fisher-Information $I(y, x, \theta)$ einer Realisation $y = (y_1, \dots, y_n)^T$ die Darstellung

$$I_{i,j}(y, x, \theta) = \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} + \frac{1}{2} \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_j}) \quad i, j = 1, \dots, k. \quad (3.2.4)$$

Beweis

Wir prüfen zunächst die Gültigkeit der Regularitätsbedingung (3.2.2). Die Dichte der Normalverteilung ist integrierbar über y , stetig differenzierbar nach θ und diese Ableitung ist für alle $y \in \mathbb{R}^n$ beschränkt. Weiterhin ist der Träger unabhängig von θ . Daher kann hier Differentiation und Integration vertauscht werden und es gilt:

$$\int_{\mathbb{R}^n} \frac{\partial^2}{\partial \theta^2} l(y, x, \theta) dy = \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}^n} l(y, x, \theta) dy = \frac{\partial^2}{\partial \theta^2} 1 = 0$$

Die Regularitätsbedingung ist also erfüllt und wir verwenden die Darstellung

$$I_{i,j}(y, x, \theta) = -E\left(\frac{\partial^2 \ln l(y, x, \theta)}{\partial \theta_i \partial \theta_j}\right), \quad i, j = 1, \dots, k$$

Die Log-Likelihood-Funktion der Normalverteilung hat die folgende Form:

$$\ln l(y, x, \theta) = \ln\left(\frac{1}{\sqrt{(2\pi)^n |\Sigma_\theta|}} e^{-\frac{1}{2}(y-\mu_\theta)^T \Sigma_\theta^{-1} (y-\mu_\theta)}\right)$$

$$= -\frac{1}{2}(n \ln(2\pi) + \ln |\Sigma_\theta| + (y - \mu_\theta)^T \Sigma_\theta^{-1} (y - \mu_\theta))$$

Mit Hilfe der folgenden beiden Beziehungen (Searle (1982)):

$$\frac{\partial \ln |\Sigma_\theta|}{\partial \theta_i} = \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i})$$

und

$$\frac{\partial \Sigma_\theta^{-1}}{\partial \theta_i} = -\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1}$$

und der Bezeichnung $Z = (y - \mu_\theta)(y - \mu_\theta)^T$ kann die Ableitung dargestellt werden durch (vgl. Magnus und Neudecker (1999))

$$\begin{aligned} \frac{\partial \ln l(y, x, \theta)}{\partial \theta_i} &= -\frac{1}{2} \left[\text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i}) + 2 \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} (y - \mu_\theta) - (y - \mu_\theta)^T \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} (y - \mu_\theta) \right] \\ &= -\frac{1}{2} \left[\text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} (\Sigma_\theta - Z)) + 2 \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} (y - \mu_\theta) \right]. \end{aligned}$$

Für die zweite Ableitung gilt:

$$\begin{aligned} \frac{\partial^2 \ln l(y, x, \theta)}{\partial \theta_i \partial \theta_j} &= -\frac{1}{2} \left[\text{tr} \left(\frac{\partial \Sigma_\theta^{-1}}{\partial \theta_j} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} (\Sigma_\theta - Z) \right) \right. \\ &\quad + \text{tr} \left(\Sigma_\theta^{-1} \frac{\partial^2 \Sigma_\theta}{\partial \theta_i \partial \theta_j} \Sigma_\theta^{-1} (\Sigma_\theta - Z) \right) \\ &\quad + \text{tr} \left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \frac{\partial \Sigma_\theta^{-1}}{\partial \theta_j} (\Sigma_\theta - Z) \right) \\ &\quad + \text{tr} \left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_j} \right) \\ &\quad + 2 \text{tr} \left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} (y - \mu_\theta)^T \right) \\ &\quad + 2 \frac{\partial^2 \mu_\theta^T}{\partial \theta_i \partial \theta_j} \Sigma_\theta^{-1} (y - \mu_\theta) \\ &\quad \left. + 2 \frac{\partial \mu_\theta^T}{\partial \theta_i} \frac{\partial \Sigma_\theta^{-1}}{\partial \theta_j} (y - \mu_\theta) + 2 \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} \right]. \end{aligned}$$

Für die zentralen Momente der Normalverteilung ist aber bekannt, das $E(y - \mu_\theta) = 0$ und $E(\Sigma_\theta - Z) = 0$. Damit ergeben sich die Komponenten der Fisher-Information folgendermaßen:

$$I_{i,j}(y, x, \theta) = -E \left(\frac{\partial^2 \ln l(y, x, \theta)}{\partial \theta_i \partial \theta_j} \right)$$

$$= \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} + \frac{1}{2} \text{tr} \left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_j} \right) \quad i, j = 1, \dots, k$$

□

Die komplette $k \times k$ Informationsmatrix hat somit die folgende Struktur:

$$I(y, x, \theta) = \frac{\partial \mu_\theta^T}{\partial \theta} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta} + \frac{1}{2} \begin{pmatrix} \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_1} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_1}) & \dots & \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_1} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_k}) \\ \vdots & & \vdots \\ \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_k} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_1}) & \dots & \text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_k} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_k}) \end{pmatrix} \quad (3.2.5)$$

Im Folgenden betrachten wir einige Spezialfälle.

Unkorrelierte Beobachtungen

Unter den Annahme unkorrelierter Beobachtungen, d.h.

$$\Sigma_\theta = \text{diag}(\sigma_\theta^2), \quad \sigma_\theta^2 = (\sigma_1^2, \dots, \sigma_n^2),$$

entspricht die Gesamtinformation einer Messreihe unter Versuchsbedingungen $x = (x_1, \dots, x_n)^T$ der Summe der Einzelinformationen:

$$\begin{aligned} I_{i,j}(y, x, \theta) &= \sum_{l=1}^n \left(\frac{1}{\sigma_l^2} \frac{\partial \mu_l}{\partial \theta_i} \frac{\partial \mu_l}{\partial \theta_j} + \frac{1}{2\sigma_l^4} \frac{\partial \sigma_l^2}{\partial \theta_i} \frac{\partial \sigma_l^2}{\partial \theta_j} \right) \quad i, j = 1, \dots, k \\ &= \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} + \frac{1}{2} \frac{\partial \sigma_\theta^2}{\partial \theta_i} \Sigma_\theta^{-2} \frac{\partial \sigma_\theta^2}{\partial \theta_j} \quad i, j = 1, \dots, k \end{aligned}$$

mit $\Sigma_\theta^{-1} = \text{diag}(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2})$.

Die gesamte Informationsmatrix hat dann die folgende Darstellung:

$$I(y, x, \theta) = \frac{\partial \mu_\theta^T}{\partial \theta} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta} + \frac{1}{2} \frac{\partial \sigma_\theta^2}{\partial \theta} \Sigma_\theta^{-2} \frac{\partial \sigma_\theta^2}{\partial \theta} \quad (3.2.6)$$

Varianz als Funktion des Erwartungswerts

Theorem 3.6

Sei $Y = (Y_1, \dots, Y_n)^T$ normalverteilt mit Erwartungswert $\mu_\theta = (\mu_1, \dots, \mu_n)^T$, abhängig von Parametern $\theta = (\theta_1, \dots, \theta_k)^T \in \Theta$. Sei weiterhin die Kovarianz Σ_θ von Y eine Diagonalmatrix

gegeben durch eine differenzierbare Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ und

$$\Sigma_\theta := \text{diag}(g(\mu_1)^2, \dots, g(\mu_n)^2). \quad (3.2.7)$$

Dann gilt:

$$I_{i,j}(y, x, \theta) = \frac{\partial \mu_\theta^T}{\partial \theta} \tilde{\Sigma}_\theta \frac{\partial \mu_\theta}{\partial \theta} \quad (3.2.8)$$

$$\text{mit } \tilde{\Sigma}_\theta = \text{diag}\left(\frac{g(\mu_1)^2 + 2\frac{\partial g(\mu_1)}{\partial \mu_1}}{g(\mu_1)^4}, \dots, \frac{g(\mu_n)^2 + 2\frac{\partial g(\mu_n)}{\partial \mu_n}}{g(\mu_n)^4}\right).$$

Beweis:

Es gilt:

$$\begin{aligned} \frac{\partial \sigma_r^2}{\partial \theta_i} &= 2 \frac{\partial g(\mu_r)}{\partial \mu_r} \frac{\partial \mu_r}{\partial \theta_i} \quad r = 1, \dots, n \\ I(y, x, \theta) &= \frac{\partial \mu_\theta^T}{\partial \theta} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta} + \frac{1}{2} \frac{\partial \sigma_\theta^{2T}}{\partial \theta} \Sigma_\theta^{-2} \frac{\partial \sigma_\theta^2}{\partial \theta} \\ &= \frac{\partial \mu_\theta^T}{\partial \theta} \left(\Sigma_\theta^{-1} + 2 \text{diag}\left(\frac{\partial g(\mu_1)}{\partial \mu_1}, \dots, \frac{\partial g(\mu_n)}{\partial \mu_n}\right) \Sigma_\theta^{-2} \right) \frac{\partial \mu_\theta}{\partial \theta} \\ &= \frac{\partial \mu_\theta^T}{\partial \theta} \tilde{\Sigma}_\theta \frac{\partial \mu_\theta}{\partial \theta} \quad \square \end{aligned}$$

3.2.2 Fisher-Information bei lognormalverteilten Fehlern

Theorem 3.7

Sei $Z = (Z_1, \dots, Z_n)^T$ normalverteilt mit Erwartungswert $\mu_\theta = (\mu_1, \dots, \mu_n)$ und Kovarianzmatrix Σ_θ und $Y = (Y_1, \dots, Y_n)^T$ lognormalverteilt mit identischen Parametern μ_θ und Σ_θ . Dann entspricht bezüglich θ die Fisher-Informationsmatrix $I(y, \theta)$ einer Realisation $y = (y_1, \dots, y_n)^T$ von Y der Informationsmatrix $I(z, \theta)$ einer Realisation $z = (z_1, \dots, z_n)^T$ von Z , d.h.

$$I_{i,j}(y, \theta) = I_{i,j}(z, \theta) = \frac{\partial \mu_\theta^T}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \mu_\theta}{\partial \theta_j} + \frac{1}{2} \text{tr}\left(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_j}\right) \quad i, j = 1, \dots, k \quad (3.2.9)$$

Beweis:

Die log Likelihood-Funktion einer Realisation $y = (y_1, \dots, y_n)^T$ der Lognormalverteilung mit

Parametern $\mu_\theta = (\mu_1, \dots, \mu_n)^T$ und $\Sigma_\theta \in \mathbb{R}^{n \times n}$ ist:

$$\begin{aligned} \ln l(y, \theta) &= \ln\left(\frac{1}{\sqrt{(2\pi)^n |\Sigma_\theta| \prod_{i=1}^n y_i}} e^{-\frac{1}{2}(\ln y - \mu_\theta)^T \Sigma_\theta^{-1} (\ln y - \mu_\theta)}\right) \\ &= -\frac{1}{2} \left[n \ln 2\pi + \sum_{i=1}^n \ln y_i + \ln |\Sigma_\theta| + (\ln y - \mu_\theta)^T \Sigma_\theta^{-1} (\ln y - \mu_\theta) \right] \end{aligned}$$

Die erste Ableitung ist:

$$\frac{\partial \ln l(y, \theta)}{\partial \theta_i} = -\frac{1}{2} \left[\text{tr}(\Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i}) + 2 \frac{\partial \mu_\theta}{\partial \theta_i} \Sigma_\theta^{-1} (\ln y - \mu_\theta) - (\ln y - \mu_\theta)^T \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_i} \Sigma_\theta^{-1} (\ln y - \mu_\theta) \right]$$

Diese Darstellung ebenso wie die entsprechende Darstellung für die zweite Ableitung unterscheidet sich von den Darstellungen der Normalverteilung nur dadurch, dass y durch $\ln(y)$ ersetzt wurde. Es gilt aber:

Ist Y lognormalverteilt mit Parametern μ und Σ so ist $\ln Y$ normalverteilt mit identischen Parametern μ und Σ .

Daraus folgt die Behauptung. \square

3.3 Exakte und approximative Designs

Offenbar hängt die Varianz aller behandelten Schätzverfahren von den Funktionen $\tilde{f}(x, \theta)$ und damit von den Versuchsbedingungen $x = (x_1, \dots, x_n)^T$ ab. Grundidee der optimalen Versuchsplanung ist es, diese Versuchsbedingungen so zu wählen, dass bestimmte Funktionale der Varianzmatrix der Parameterschätzung möglichst klein werden (siehe Abschnitt 3.6).

Eine solche Aufteilung der n verfügbaren Beobachtungen auf p verschiedene Versuchsbedingungen wird als *Design* oder *Versuchsplan* bezeichnet.

Definition 3.8

Ein Wahrscheinlichkeitsmaß ξ auf $\mathcal{X} \subset \mathbb{R}$ mit endlichem Träger, p Trägerpunkten und Wahrscheinlichkeitsmassen $\xi(x_i) > 0$, $i = 1, \dots, p$, die Vielfache von $\frac{1}{n}$ sind, heißt „Exaktes Design zum Stichprobenumfang n “. Ein beliebiges Wahrscheinlichkeitsmaß auf \mathcal{X} mit endlichem Träger

heißt „Approximatives Design“.

Die Mengen aller exakten und approximativen Designs werden bezeichnet mit

$$\Xi_n = \{\xi \mid \xi \text{ ist exaktes Design zum Stichprobenumfang } n\} \quad (3.3.1)$$

und

$$\Xi = \{\xi \mid \xi \text{ ist approximatives Design}\} \supseteq \Xi_n. \quad (3.3.2)$$

Eine tatsächliche Aufteilung von n Beobachtungen auf Versuchsbedingungen $x = (x_1, \dots, x_n)^T$ sei als „ n -Punkt Design“ bezeichnet.

In beiden Fällen kann ein n -Punkt Design aus den Wahrscheinlichkeitsmaßen durch Multiplikation der Wahrscheinlichkeitsmassen $\xi(x)$ mit n bestimmt werden. Im Falle approximativer Designs führt dies nicht immer zu ganzzahligen Messhäufigkeiten, so dass geeignete Rundungsverfahren notwendig sind (siehe Pukelsheim und Rieder (1992)).

Die Menge Ξ_n ist in der Regel nicht konvex. Da der entstehende Verlust an Schätzgenauigkeit durch Rundungsverfahren gering ist, wird üblicherweise über die Menge der approximativen Versuchspläne Ξ optimiert.

Beispiel

Das approximative Design

$$\xi = \begin{pmatrix} 0 & 2 \\ 0.4 & 0.6 \end{pmatrix}.$$

weist den Untersucher an, 40% der Beobachtungen mit Versuchsbedingung 0 und 60% mit Versuchsbedingung 2 durchzuführen. Bei $n = 10$ Patienten heißt das, 4 werden unter Bedingung 0 und 6 unter 2 behandelt. In diesem Fall ist $\xi \in \Xi_{10}$.

3.4 Informationsmatrizen

Optimale Designs haben zum Ziel, Funktionale der Varianzmatrix zu minimieren. Ein äquivalenter und technisch einfacherer Ansatz ist es, angepasste Funktionale der Inverse der

Varianzmatrix zu minimieren/maximieren. Solche Matrizen werden als *Informationsmatrizen* bezeichnet. Mit der Fisher-Information wurde bereits ein Vertreter dieser Klasse eingeführt.

3.4.1 Kleinste-Quadrate-Schätzer

Definition 3.9

Gegeben sei das Modell (2.2.2) mit Versuchsbedingungen $x = (x_1, \dots, x_n)^T$. Sei $\hat{\theta}$ ein NWLS-Schätzer für $\theta \in \Theta$. Dann bezeichne

$$M(x) = \tilde{F}V^{-1}\tilde{F}^T \quad (3.4.1)$$

die Informationsmatrix der Versuchsbedingungen x für den NWLS-Schätzer.

Bezeichnet man die Elemente der Inverse der Matrix V mit v_{ij} , $i, j = 1, \dots, n$, so gilt für den Spezialfall unkorrelierter Beobachtungen

$$M(x) = \tilde{F}V^{-1}\tilde{F}^T = \sum_{i=1}^n \tilde{f}(x_i, \theta)\tilde{f}(x_i, \theta)^T v_{ii}, \quad (3.4.2)$$

d.h. die Gesamtinformationsmatrix einer Messreihe entspricht der Summe der Einzelinformationen.

Definition 3.10

Gegeben sei das Modell (2.2.2) mit unkorrelierten Beobachtungen und ein Design $\xi \in \Xi$ mit p Trägerpunkten. Dann bezeichne

$$M(\xi) = \sum_{i=1}^p \tilde{f}(x_i, \theta)\tilde{f}(x_i, \theta)^T v_{ii}\xi(x_i) \quad (3.4.3)$$

die (normierte) Informationsmatrix des Designs ξ für den NWLS-Schätzer. Die konvexe Hülle dieser Informationsmatrizen wird mit

$$\mathcal{M} = \text{conv} \{M(\xi) | \xi \in \Xi\} \subset \mathbb{R}^{k \times k} \quad (3.4.4)$$

bezeichnet.

Im Fall korrelierter Beobachtungen, d.h. beliebiger Kovarianzmatrix V , kann die Gesamtinformation einer Messreihe basierend auf einem n -Punkt Design nicht mehr als Summe der Einzelinformationen dargestellt werden. Demnach kann auch die Informationsmatrix eines approximativen Versuchsplans nicht allgemein dargestellt werden, und die Anwendung gängiger Techniken der Versuchsplanung (vgl. Kapitel 4 und 5) ist erschwert.

Populationsmodelle

Bei Populationsmodellen (Modell 2.3.1) kann nur im Spezialfall einer einzigen Beobachtung pro Patient davon ausgegangen werden, dass alle Beobachtungen des Vektors Y stochastisch unabhängig sind. Bei mehr als einer Beobachtung pro Individuum sorgt der gemeinsame zufällige Effekt dafür, dass Beobachtungen am selben Individuum korreliert sind, auch dann wenn die reinen Störgrößen unkorreliert sind. Auch in dieser Situation kann daher die Informationsmatrix eines approximativen Designs über \mathcal{X} nicht allgemein dargestellt werden. Es kann jedoch davon ausgegangen werden, dass Messungen an unterschiedlichen Individuen miteinander unkorreliert sind. Angenommen, pro Individuum stünden m Messungen zur Verfügung ($m > 1$), so ist es als Ausweg in dieser Situation möglich, die Versuchsbedingungen aller m Messungen am Individuum i ($i = 1, \dots, N$) zu einem Vektor $x_i = (x_{i1}, \dots, x_{im})^T$ zusammenzufassen, und die Menge dieser Vektoren \mathcal{X}^m als neuen Versuchsbereich zu definieren. Die eigentliche Optimierung beinhaltet dann, den verfügbaren N Patienten optimale m -Punkt Vektoren an Messbedingungen zuzuweisen (m -Punkt-Messsätze).

Notwendig für diese Betrachtung ist es, die Kovarianzmatrix der Beobachtungen $Y = (Y_{11}, \dots, Y_{1m}, Y_{21}, \dots, Y_{Nm})^T$ genauer darzustellen. Wir bezeichnen $\tilde{F}_i = (\tilde{f}(x_{i1}, \theta), \dots, \tilde{f}(x_{im}, \theta))$ als Designmatrix des i -ten Individuums, $W_i \in \mathbb{R}^{m \times m}$ als Kovarianz der Störgrößen des i -ten Individuums sowie

$$V_i = \tilde{F}_i^T \Omega \tilde{F}_i + W_i \in \mathbb{R}^{m \times m} \quad (3.4.5)$$

als positiv semidefinite Kovarianzmatrix der m Beobachtungen am Patienten i , $i = 1, \dots, n$.

Dann ist

$$V = \text{diag}(V_1, \dots, V_n) \in \mathbb{R}^{Nm \times Nm}$$

die gesamte Kovarianzmatrix von Y . Mit diesen Darstellungen ist es möglich, Informationsmatrizen für approximative Designs in Populationsmodelle zu definieren.

Definition 3.11

Die Informationsmatrix des NWLS-Schätzers für eine Beobachtungsreihe unter Versuchsbedingungen $x = (x_1, \dots, x_N)^T$ im Populationsmodell ist gegeben durch

$$M(x) = \sum_{i=1}^N \tilde{F}_i V_i^{-1} \tilde{F}_i^T. \quad (3.4.6)$$

Die Informationsmatrix eines approximativen Designs über \mathcal{X}^m mit p Trägerpunkten ist gegeben durch

$$M(\xi) = \sum_{i=1}^p \tilde{F}_i V_i^{-1} \tilde{F}_i^T \xi(x_i). \quad (3.4.7)$$

Die konvexe Hülle dieser Informationsmatrizen wird mit

$$\mathcal{M} = \text{conv} \{M(\xi) | \xi \in \Xi\} \subset \mathbb{R}^{k \times k} \quad (3.4.8)$$

bezeichnet.

Das folgende Lemma erlaubt eine alternative Darstellung dieser Informationsmatrix in einer Form ähnlicher zu der Form bei unkorrelierten Darstellungen (vgl. Holland-Letz et al. (2009)).

Lemma 3.12

Eine Informationsmatrix der Form $\tilde{F}_i V_i^{-1} \tilde{F}_i^T$ kann dargestellt werden als

$$\tilde{F}_i V_i^{-1} \tilde{F}_i^T = \sum_{\ell=1}^m \tilde{f}_\ell(x_i, \theta) \tilde{f}_\ell(x_i, \theta)^T, \quad (3.4.9)$$

wobei die Funktionen $\tilde{f}_\ell : \mathcal{X}^m \times \Theta \rightarrow \mathbb{R}^k$ definiert sind als die Spalten der $k \times m$ -Matrix $F_i^* = \tilde{F}_i V_i^{-0.5}$.

Beweis:

Bezeichne $V_i^{-1} = UDU^T$ die Eigenwertzerlegung der (positiv semidefiniten) Matrix V_i^{-1} , mit U die Matrix der Eigenvektoren und $D = \text{diag}(\lambda_1, \dots, \lambda_m)$ die Diagonalmatrix der Eigenwerte von V_i^{-1} . Dann ist

$$V_i^{-0.5} = U \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m}) U^T$$

die *Wurzel* der Matrix V_i^{-1} , d.h. $V_i^{-0.5}V_i^{-0.5} = V_i^{-1}$.

Dann gilt:

$$\tilde{F}_i V_i^{-1} \tilde{F}_i^T = \tilde{F}_i V_i^{-0.5} V_i^{-0.5} \tilde{F}_i^T$$

Schreibt man

$$F_i^* := \tilde{F}_i V_i^{-0.5} \in \mathbb{R}^{k \times m}$$

und bezeichne $\tilde{f}_\ell(x_i, \theta)$ die ℓ -te Spalte von F_i^* , so gilt:

$$\tilde{F}_i V_i^{-1} \tilde{F}_i^T = F_i^* (F_i^*)^T = \sum_{i=1}^m \tilde{f}_\ell(x_i, \theta) \tilde{f}_\ell(x_i, \theta)^T \quad \square$$

Mittels Lemma 3.12 kann die Gesamtinformation eines n -Punkt Designs für m Beobachtungen an je N Individuen dargestellt werden als

$$M(x) = \sum_{i=1}^N \sum_{\ell=1}^m \tilde{f}_\ell(x_i, \theta) \tilde{f}_\ell(x_i, \theta)^T, \quad (3.4.10)$$

und entsprechend ein approximatives Design mit p Trägerpunkten als

$$M(\xi) = \sum_{i=1}^p \sum_{\ell=1}^m \tilde{f}_\ell(x_i, \theta) \tilde{f}_\ell(x_i, \theta)^T \xi(x_i). \quad (3.4.11)$$

3.4.2 Maximum Likelihood Schätzer

Unkorrelierte Beobachtungen

Im Gegensatz zum Kleinste Quadrate-Schätzer ist der Maximum-Likelihood-Schätzer auch in der Lage, aus der Varianz der Beobachtungen Information über die Parameter zu gewinnen, falls diese einen Einfluss auf die Varianz ausüben. Die Informationsmatrix im Modell (2.2.2) setzt sich daher aus zwei Komponenten zusammen (vgl. (3.2.5) und (3.2.6), mit $\mu_\theta = \eta(x, \theta)$)

und $\Sigma_\theta = V$).

Definition 3.13

Gegeben sei das Modell 2.2.2 mit unkorrelierten Beobachtungen, normalverteilten Störgrößen und Versuchsbedingungen $x = (x_1, \dots, x_n)^T$. Sei $\hat{\theta}$ ein ML-Schätzer für $\theta \in \Theta$. Dann bezeichne

$$M(x) = \sum_{i=1}^n (\tilde{f}_1(x_i, \theta) \tilde{f}_1(x_i, \theta)^T + \tilde{f}_2(x_i, \theta) \tilde{f}_2(x_i, \theta)^T) \quad (3.4.12)$$

die Informationsmatrix der Versuchsbedingungen x für den Maximum-Likelihood-Schätzer-Schätzer, mit

$$\tilde{f}_1(x, \theta) := \tilde{f}(x, \theta) \frac{1}{\sigma(x, \theta)} \quad \text{und} \quad \tilde{f}_2(x, \theta) := \frac{1}{\sqrt{2}} \frac{\partial \sigma^2(x, \theta)}{\partial \theta} \frac{1}{\sigma^2(x, \theta)}. \quad (3.4.13)$$

Dabei sei $\sigma^2(x_i, \theta) = \text{Var}(Y_i)$ wie in (2.2.4).

Alternativ kann diese Informationsmatrix als

$$M(x) = \sum_{i=1}^n \sum_{j=1}^2 \tilde{f}_j(x_i, \theta) \tilde{f}_j(x_i, \theta)^T, \quad (3.4.14)$$

dargestellt werden.

Im Falle parameterunabhängiger Varianz ist der zweite Teil der Summe gleich Null.

Definition 3.14

Gegeben sei das Modell 2.2.2 mit unkorrelierten Beobachtungen und ein Design $\xi \in \Xi$. Dann bezeichne

$$M(\xi) = \sum_{i=1}^p \sum_{j=1}^2 \tilde{f}_j(x_i, \theta) \tilde{f}_j(x_i, \theta)^T \xi(x_i) \quad (3.4.15)$$

die (normierte) Informationsmatrix des Designs ξ für den Maximum-Likelihood-Schätzer.

Die konvexe Hülle dieser Informationsmatrizen wird mit

$$\mathcal{M} = \text{conv} \{M(\xi) | \xi \in \Xi\} \subset \mathbb{R}^{m \times m} \quad (3.4.16)$$

bezeichnet.

Korrelierte Beobachtungen

Für Modelle mit korrelierten Beobachtungen allgemeiner Struktur ist die Information einer Versuchsreihe gegeben durch (3.2.5). Die Darstellung der Informationsmatrix eines approximativen Versuchsplans ist auch hier nicht allgemein möglich.

Im Spezialfall parameterunabhängiger Varianz entspricht die Gesamtinformationsmatrix des ML-Schätzers der des NWLS-Schätzers (vgl. 3.1.10).

Im Spezialfall der Populationsmodelle mit Unabhängigkeit zwischen Beobachtungen verschiedener Patienten kann durch Definition eines neuen Versuchsbereiches auch hier wieder eine Darstellung approximativer Designs über den Raum der m -Punkt Messsätze \mathcal{X}^m erlangt werden. Bezeichne wieder

$$V_i = \tilde{F}_i^T \Omega \tilde{F}_i + W_i \in \mathbb{R}^{m \times m} \quad (3.4.17)$$

die Kovarianz der m Beobachtungen am Patienten i und \tilde{F}_i die zugehörige Designmatrix, $i = 1, \dots, N$. Die Informationsmatrix A_i aller Beobachtungen eines Patienten ist dann gegeben durch:

$$A_i = \tilde{F}_i V_i^{-1} \tilde{F}_i^T + \frac{1}{2} \left\{ \begin{array}{ccc} \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \dots & \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}) \\ \vdots & & \vdots \\ \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \dots & \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}). \end{array} \right\} \quad (3.4.18)$$

Aufgrund der Unabhängigkeit zwischen Patienten kann die Informationsmatrix eines approximativen Designs ξ über den Raum der m -Punkt Messsätze \mathcal{X}^m durch

$$M(\xi) = \sum_{i=1}^p A_i \xi(x_i) \quad (3.4.19)$$

dargestellt werden.

Die Darstellung als Summe der Form (3.4.11), analog zu Lemma 3.12, ist noch immer möglich, hier aber erheblich aufwändiger.

3.5 Informationsfunktionen

Die im vorherigen Kapitel eingeführten Informationsmatrizen beschreiben die Information eines Versuchsplans für k verschiedene Parameter. Für $k > 1$ ist es jedoch im allgemeinen nicht möglich,

die Information für alle Parameter simultan zu maximieren, d.h. es gibt für $k > 1$ kein Design ξ^* das im strengen Sinne der Loewner Halbordnung optimal ist, so dass $M(\xi^*) - M(\xi)$ positiv semidefinit für alle $\xi \in \Xi$ (Pukelsheim (1993)). Aus diesem Grund wird versucht, bestimmte Funktionale ϕ der Informationsmatrix zu maximieren. Gängige Funktionale sind

$$\ln|M(\xi)|$$

(D-Optimalität, $-\infty$ falls $|M(\xi)| = 0$) oder

$$-c^T M^{-}(\xi)c$$

für einen Vektor $c \in \mathbb{R}^k$ (c -Optimalität, $M^{-}(\xi)$ sei verallgemeinerte Inverse von $M(\xi)$). D-optimale Designs erlauben eine gleichmäßig gute Schätzung aller Parameter, während c -optimale Designs für die Schätzung einer bestimmten Linearkombination $c^T\theta$ der Parameter optimiert sind.

In einigen Fällen (speziell in nichtlinearen Modellen) hängt die Informationsmatrix und damit das optimale Design von den tatsächlichen Werten der Parameter ab. Solche Designs bezeichnet man als *lokal optimal* (Chernoff (1953)).

3.5.1 Schätzbarkeit

Im Falle D-optimaler Designs kann ein Design mit weniger als k Trägerpunkten niemals optimal sein (in diesem Fall ist die Determinante der Informationsmatrix gleich Null). Für c -Optimalität ebenso wie für allgemeinere Formulierungen von Informationsfunktionen ist dies nicht allgemein der Fall, Designs mit weniger als k Trägerpunkten sind denkbar. Im Fall von c -Optimalität ist eine Schätzung der interessierenden Linearkombination $c^T\theta$ nur möglich, falls der Vektor $c \in \mathbb{R}^k$ im Spaltenraum der Informationsmatrix liegt. Man spricht in diesem Fall davon, dass der Vektor c durch das Design *schätzbar* ist.

Definition 3.17

Eine Linearkombination $c^T\theta$ ($c \neq 0$) heißt *schätzbar* durch ein Design ξ , falls ein Vektor $a \in \mathbb{R}^k$

existiert, so dass $M(\xi)a = c$. Entsprechend ist der Ausdruck $K^T\theta$ mit $K \in \mathbb{R}^{k \times r}$ ($K \neq 0$) schätzbar, falls ein $A \in \mathbb{R}^{k \times r}$ existiert, so daß $M(\xi)A = K$.

Die Menge der $M(\xi)$ für die der Ausdruck $K^T\theta$ durch ξ schätzbar ist, wird als $\mathcal{A}(K)$ bezeichnet.

Eine hinreichende Bedingung zur Schätzbarkeit ist es, das $\text{Rang}(M(\xi)) = k$. Designs mit maximalem Wert der Informationsfunktion, die keine Schätzung erlauben, nennt man *formal optimal*.

3.5.2 ϕ^p -Optimalität

Definition 3.15

Sei K eine beliebige $k \times l$ -Matrix und $\mathcal{C} = \{C_K(\xi) | C_K(\xi) = (K^T M(\xi)^{-1} K)^{-1}, \xi \in \mathcal{A}(K)\}$. Weiterhin sei $\phi : \mathcal{C} \rightarrow \mathbb{R}$ eine Funktion mit den folgenden Eigenschaften:

1. Positive Homogenität: $\phi(\lambda C) = \lambda \phi(C) \quad \forall \quad \lambda > 0, C \in \mathcal{C}$.
2. Super-Additivität: $\phi(B + C) \geq \phi(B) + \phi(C) \quad \forall \quad B, C \in \mathcal{C}$.
3. Nichtnegativität: $\phi(C) \geq 0 \quad \forall \quad C \in \mathcal{C}$.
4. Nicht konstant.
5. Halbstetigkeit von oben: $C_n \rightarrow C, C_n \geq C \Rightarrow \phi(C_n) \rightarrow \phi(C), \quad C_n, C \in \mathcal{C}$.

Ein approximatives Design ξ heißt lokal ϕ -optimal, wenn es $\phi(C_K(\xi))$ maximiert für alle $\xi \in \Xi$. Die Funktion ϕ wird Informationsfunktion genannt.

Es existieren eine Vielzahl von möglichen Informationsfunktionen (Pázman (1986)). Die wichtigsten werden durch die folgende allgemeine Definition abgedeckt .

Definition 3.16

Ein Design ξ_p heißt lokal ϕ^p -optimal (optimal nach p -tem Matrix-Mittel, $p \in [-\infty, 1]$) falls für

alle $\xi \in \Xi$ gilt:

$$\begin{aligned}
\left[\frac{1}{k} \operatorname{tr}(C_K^p(\xi_p))\right]^{\frac{1}{p}} &\geq \left[\frac{1}{k} \operatorname{tr}(C_K^p(\xi))\right]^{\frac{1}{p}} \quad \text{falls } p \neq 0, -\infty \\
|C_K(\xi_p)|^{\frac{1}{k}} &\geq |C_K(\xi)|^{\frac{1}{k}} \quad \text{falls } p = 0 \\
\lambda_{\min}(\xi_p) &\geq \lambda_{\min}(\xi) \quad \text{falls } p = -\infty
\end{aligned} \tag{3.5.1}$$

Dazu bezeichne $\lambda_{\min}(\xi)$ den kleinsten Eigenwert der Matrix $C_K(\xi)$.

Optimalitätskriterien äquivalent zur D- und c-Optimalität sind in diesem Kriterium als Spezialfälle mit $p = 0$ und $K = I_k$ (D-Optimalität) bzw. $K = c \in \mathbb{R}^k$ (c-Optimalität) enthalten. Der Ausdruck für $p = 0$ entsteht auch als Grenzwert für $p \rightarrow 0$:

Es seien dazu $(\lambda_1, \dots, \lambda_k)$ die Eigenwerte von $C_K(\xi)$. Man erhält:

$$\begin{aligned}
\lim_{p \rightarrow 0} \ln \left[\frac{1}{k} \operatorname{tr}(C_K(\xi)^p) \right]^{\frac{1}{p}} &= \lim_{p \rightarrow 0} \frac{\ln \frac{1}{k} \sum_{i=1}^k \lambda_i^p}{p} \\
&\stackrel{\text{l'Hospital}}{=} \lim_{p \rightarrow 0} \frac{\partial \ln \frac{1}{k} \sum_{i=1}^k \lambda_i^p}{\partial p} \\
&= \lim_{p \rightarrow 0} \frac{1}{\frac{1}{k} \sum_{i=1}^k \lambda_i^p} \frac{\partial \frac{1}{k} \sum_{i=1}^k e^{p \ln \lambda_i}}{\partial p} \\
&= \lim_{p \rightarrow 0} \frac{1}{\frac{1}{k} \sum_{i=1}^k \lambda_i^p} \frac{1}{k} \sum_{i=1}^k (\ln \lambda_i) \lambda_i^p \\
&= \frac{1}{k} \sum_{i=1}^k \ln \lambda_i \\
&= \frac{1}{k} \ln \prod_{i=1}^k \lambda_i \\
&= \ln |C_K(\xi)|^{\frac{1}{k}}
\end{aligned}$$

Auf ähnliche Weise kann auch ein Grenzübergang für $p \rightarrow -\infty$ durchgeführt werden.

3.5.3 c-Optimalität für abgeleitete Kennzahlen der Pharmakokinetik

In Fragestellungen der Pharmakokinetik ist das primäre Studieninteresse häufig nicht die eigentliche Schätzung der Parameter der Konzentrationsfunktionen, sondern stattdessen die Schätzung bestimmter zusammengefasster Wirksamkeitsparameter $g(\theta)$, die von den Parametern abhängen. Die Wichtigsten solcher Kennzahlen sind die Fläche unter der Konzentrationskurve (AUC) als Maß für den Gesamteffekt einer Medikation und der Zeitpunkt x_{max} der Maximalkonzentration (Atkinson et al. (1993)) als Ausdruck für den Zeitpunkt des stärksten zu erwartenden temporären Effekts.

Wie bereits zuvor bei der Varianz für eine Schätzung der Modellparameter kann auch die Varianz der Schätzung solcher abgeleiteten Kennzahlen durch eine Taylorentwicklung erster Ordnung angenähert werden:

$$\begin{aligned} g(\theta) &\approx g(\theta_0) + \frac{\partial g(\theta_0)^T}{\partial \theta_0}(\theta - \theta_0) \\ \text{Var}(g(\theta)) &\approx \text{Var}\left(\frac{\partial g(\theta_0)^T}{\partial \theta_0}\theta\right) = \frac{\partial g(\theta_0)^T}{\partial \theta_0} \text{Var}(\theta) \frac{\partial g(\theta_0)}{\partial \theta_0} \end{aligned} \quad (3.5.2)$$

Man erhält den Schätzfehler als die Varianz der Schätzung einer Linearkombination $c^T\theta$ der Parameter. In diesem Fall sind die optimalen Designs zur Schätzung der Kennzahlen also c -optimale Designs. Die zu verwendenden c -Vektoren sind durch die ersten Ableitungen der Kennzahlen nach den Parametern gegeben.

Fläche unter der Kurve (AUC)

Die Fläche unter der Kurve läßt sich für allgemeine Kompartimentmodelle der Form 2.4.3 folgendermaßen bestimmen:

$$AUC = \int_0^\infty \sum_{i=1}^r \theta_i e^{-\theta_{i+r}x} dx = \sum_{i=1}^r -\frac{\theta_i}{\theta_{i+r}} e^{-\theta_{i+r}x} \Big|_0^\infty = \sum_{i=1}^r \frac{\theta_i}{\theta_{i+r}} \quad (3.5.3)$$

Damit ergeben sich für die in Abschnitt 2.4 besprochenen Spezialfälle:

$$AUC(Evasion) = \frac{\theta_1}{\theta_2} \quad (3.5.4)$$

$$c_{Evasion} = \frac{1}{\theta_2} \left(1, -\frac{\theta_1}{\theta_2}\right) \quad (3.5.5)$$

$$AUC(Bateman) = \frac{\theta_1}{\theta_2} - \frac{\theta_1}{\theta_3} \quad (3.5.6)$$

$$c_{Bateman} = \left(\frac{1}{\theta_2} - \frac{1}{\theta_3}, -\frac{\theta_1}{\theta_2^2}, \frac{\theta_1}{\theta_3^2}\right) \quad (3.5.7)$$

Zeitpunkt der Maximalkonzentration

Der Maximalzeitpunkt x_{max} lässt sich für allgemeine Kompartimentmodelle nicht analytisch bestimmen. Für die Spezialfälle gilt:

$$x_{max}(Evasion) = 0 \quad (3.5.8)$$

Der Zeitpunkt der maximalen Konzentration ist bei der einfachen Evasion bekannt und konstant 0 und muss nicht geschätzt werden. Entsprechend ist auch kein c -Vektor notwendig.

Für die Bateman-Funktion erhält man durch direkte Maximierung:

$$x_{max}(Bateman) = \frac{\ln \theta_2 - \ln \theta_3}{\theta_2 - \theta_3} \quad (3.5.9)$$

und damit

$$c_{Bateman} = \left(0, \frac{1}{\theta_2(\theta_2 - \theta_3)} - \frac{\ln \theta_2 - \ln \theta_3}{(\theta_2 - \theta_3)^2}, -\frac{1}{\theta_3(\theta_2 - \theta_3)} + \frac{\ln \theta_2 - \ln \theta_3}{(\theta_2 - \theta_3)^2}\right). \quad (3.5.10)$$

Kapitel 4

Äquivalenzsätze

Äquivalenzsätze liefern nachprüfbar Bedingungen für optimale Designs, d.h. mit ihrer Hilfe kann bestätigt werden, dass ein vorliegendes Design tatsächlich optimal ist. Die erste Variante dieses Types bezog sich auf D-Optimalität (Kiefer und Wolfowitz (1960)). Hier wird eine allgemeinere Variante von Pukelsheim (1993) verwendet.

4.1 Der allgemeine Äquivalenzsatz für ϕ^p -Optimalität

Theorem 4.1

Sei $K \in \mathbb{R}^{k \times l}$ und $M(\xi) \in \mathcal{M} \cap \mathcal{A}(K)$. Ein Design ξ ist lokal Φ^p -optimal in \mathcal{M} für $K^T \theta$ genau dann, wenn ein $G \in M(\xi)^-$ existiert, so daß

$$\langle A, GK(K^T M(\xi)^- K)^{-1-p} K^T G^T \rangle \leq \text{tr}(K^T M(\xi)^- K)^{-p} \quad \forall A \in \mathcal{M}. \quad (\text{Normalenungleichung})$$

In dieser Ungleichung gilt das Gleichheitszeichen für jede lokal ϕ^p -optimale Informationsmatrix.

Beweis:

Satz 7.19 in Pukelsheim (1993).

Korollar 4.2

Sei $c \in \mathbb{R}^k$ und $M(\xi) \in \mathcal{M} \cap \mathcal{A}(c)$. Ein Design ξ ist lokal c -optimal für $c^T \theta$ in \mathcal{M} genau dann,

wenn ein $G \in M(\xi)^-$ existiert, so dass

$$\langle A, Gcc^T G^T \rangle \leq c^T M(\xi)^- c \quad \forall A \in \mathcal{M} \quad (\text{Normalenungleichung}) \quad (4.1.1)$$

In dieser Ungleichung gilt das Gleichheitszeichen für jede lokal c -optimale Informationsmatrix.

Beweis:

Theorem 4.1 mit $p = 0$ und $K = c$.

Korollar 4.3

Bezeichne $M(x) = \sum_{\ell=1}^m \tilde{f}_\ell(x, \theta) \tilde{f}_\ell(x, \theta)^T$ die Informationsmatrix einer einzelnen Beobachtung x .

Dann gilt:

Ein Design ξ mit $M(\xi) \in \mathcal{M} \cap \mathcal{A}(c)$ ist lokal c -optimal für $c^T \theta$ in \mathcal{M} genau dann, wenn ein $G \in M(\xi)^-$ existiert, so dass

$$\sum_{\ell=1}^m \frac{(\tilde{f}_\ell(x, \theta)^T G c)^2}{c^T M(\xi)^- c} \leq 1 \quad \forall x \in \mathcal{X}. \quad (\text{Normalenungleichung}) \quad (4.1.2)$$

In dieser Ungleichung gilt das Gleichheitszeichen für alle Trägerpunkte von ξ .

Beweis:

Einsetzen in Korollar 4.2 liefert:

$$\langle A, Gcc^T G^T \rangle \leq c^T M(\xi)^- c \quad \forall A \in \mathcal{M}$$

$$\Leftrightarrow \langle M(x), Gcc^T G^T \rangle \leq c^T M(\xi)^- c \quad \forall x \in \mathcal{X}$$

$$\Leftrightarrow \text{tr}(M(x)Gcc^T G^T) \leq c^T M(\xi)^- c \quad \forall x \in \mathcal{X}$$

$$\Leftrightarrow \sum_{\ell=1}^m \text{tr}(\tilde{f}_\ell(x, \theta) \tilde{f}_\ell(x, \theta)^T Gcc^T G^T) \leq c^T M(\xi)^- c \quad \forall x \in \mathcal{X}$$

$$\Leftrightarrow \sum_{\ell=1}^m (\tilde{f}_\ell(x, \theta)^T G c)^2 \leq c^T M(\xi)^- c \quad \forall x \in \mathcal{X}$$

$$\Leftrightarrow \sum_{\ell=1}^m \frac{(\tilde{f}_\ell(x, \theta)^T Gc)^2}{c^T M(\xi)^{-1} c} \leq 1 \quad \forall x \in \mathcal{X}$$

Für jeden Trägerpunkt muss in dieser Gleichung \leq gelten, für die gewichtete Summe aber nach Korollar 4.2 Gleichheit. Daraus folgt die Gleichheit für jeden Trägerpunkt. \square

Definiert man die Konstante $\gamma^2 = (c^T M^{-1} c)^{-1}$ und den Vektor $d = \gamma Gc \in \mathbb{R}^k$, so ist Darstellung (4.1.2) äquivalent zu

$$\sum_{\ell=1}^r (d^T f_\ell(x))^2 \leq 1 \quad \forall x \in \mathcal{X}. \quad (4.1.3)$$

4.2 Beispiele

Als Beispiel betrachten wir eine ML-Schätzung im Einkompartiment-Modell mit intravenöser Verabreichung (Evasionsmodell) mit fixen Effekten in der Form 2.1.3

$$Y_i = \theta_1 e^{-\theta_2 x_i} + \epsilon_i, \quad i = 1, \dots, n,$$

mit den Parametern $\theta_1 = 46, \theta_2 = 1.7, \epsilon_i \sim \mathcal{N}(0, 1)$ unabhängig identisch, $c = (1, 10)^T$ und dem Versuchsbereich $\mathcal{X} = [0, 4]$. Die Funktion \tilde{f} ist damit $\tilde{f}(x_i, \theta) = (e^{-\theta_2 x_i}, -x_i \theta_1 e^{-\theta_2 x_i})^T$.

Der zugehörige Äquivalenzsatz ist durch Korollar 4.3 mit $m = 1$ gegeben.

Ein Kandidat für ein c -optimales Design in dieser Situation ist das Design

$$\xi = \begin{pmatrix} 0 & 0.75 \\ 0.554 & 0.446 \end{pmatrix}.$$

Abbildung 4.1 stellt die linke Seite von Gleichung 4.1.2 dar, geplottet gegen alle $x \in \mathcal{X}$ auf einem engen Raster. Man erkennt, dass diese Terme für alle $x \in \mathcal{X}$ kleiner eins sind, mit Gleichheit an den Trägerpunkten 0 und 0.75. Das vorgeschlagene Design ist also in der Tat optimal. Da das Design ξ zwei Trägerpunkte besitzt, ist $M(\xi) \in \mathcal{A}(c)$, d.h. die Linearkombination $c^T \theta$ ist durch ξ schätzbar.

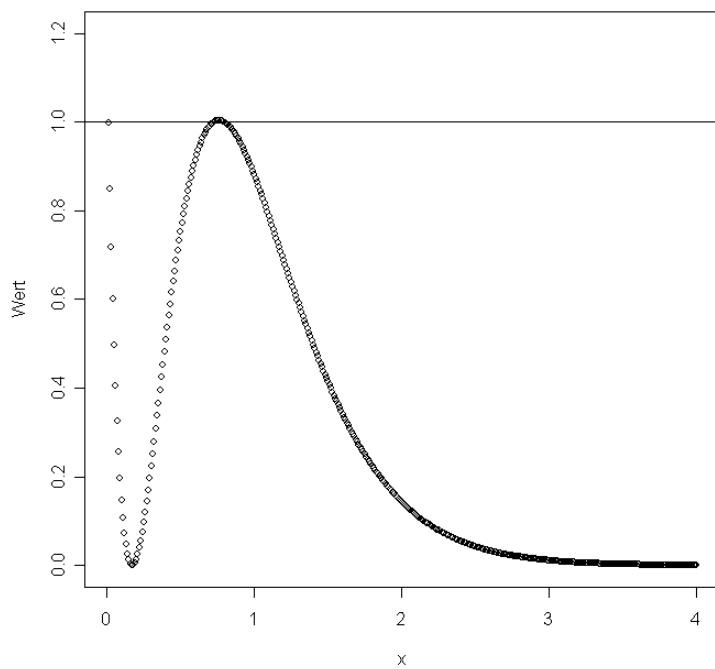


Abbildung 4.1: *Der Äquivalenzsatz aus Korollar 4.3 für ein optimales Design im Evasionsmodell mit fixen Effekten*

Zur Veranschaulichung der Eigenschaften des Äquivalenzsatzes bei nichtoptimalen Designs enthält Abbildung 4.2 analoge Darstellungen für das selbe Design mit optimalen Trägerpunkten aber (nichtoptimalen) Gewichten 0.3 und 0.7 (linke Seite) und für nichtoptimale Trägerpunkte 0 und 0.9 mit den ursprünglichen Gewichten 0.554 und 0.446.

Wie erwartet sind die Bedingungen von Korollar 4.3 in beiden Fällen nicht erfüllt.

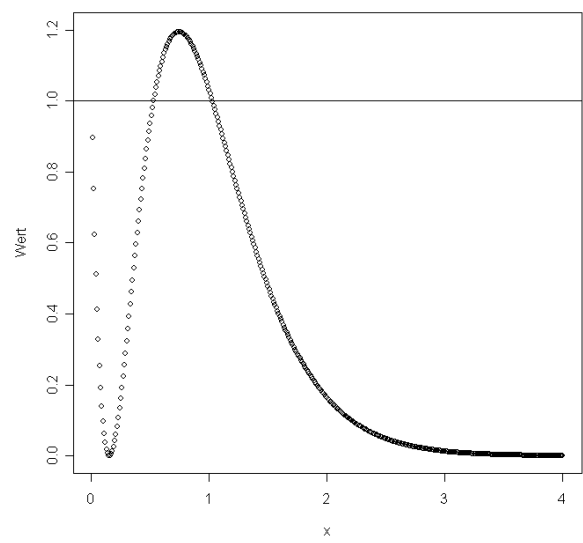
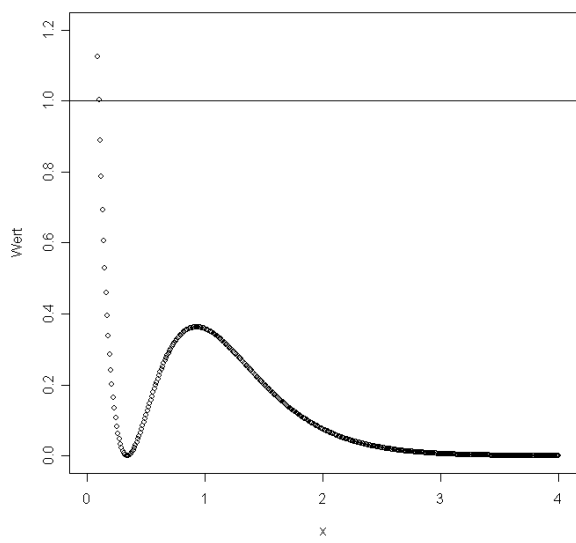


Abbildung 4.2: Der Äquivalenzsatz aus Korollar 4.3 für ein Design mit nichtoptimalen Gewichten (links) bzw. nichtoptimalen Trägerpunkten (rechts) im Evasionsmodell mit fixen Effekten.

Kapitel 5

Elfving-Sätze

Elfving-Sätze ermöglichen es, die Information aller potentiellen Ausprägungen der Designvariablen geometrisch darzustellen. Auf diese Weise ist ein graphisches Veranschaulichen und in manchen Fällen Herleiten c -optimaler Versuchspläne möglich. Ein grundlegender Fall dieses Prinzips wurde in Elfving (1952) eingeführt. In diesem Abschnitt wird eine allgemeine Version dieses Ansatzes formuliert, der das ursprüngliche Resultat von Elfving (1952) als Spezialfall enthält. Die für verschiedene Problemstellungen der Pharmakokinetik notwendigen Varianten ergeben sich als Spezialfälle des allgemeinen Theorems.

5.1 Satz von Elfving

Definition 5.1

Gegeben sei das Modell mit zufälligen Effekten (2.2.2). Die Menge $R = \text{conv}(\{\tilde{f}(x, \theta) : x \in \mathcal{X}\} \cup \{-\tilde{f}(x, \theta) : x \in \mathcal{X}\})$ wird als *Elfving-Raum* bezeichnet.

Theorem 5.2 (Elfving, 1952)

Im Modell (2.2.2) ist ein Design ξ mit Trägerpunkten x_i , Gewichten $\xi(x_i)$, $i = 1, \dots, p$ und Informationsmatrix $M(\xi) = \sum_{i=1}^p \tilde{f}(x_i, \theta) \tilde{f}(x_i, \theta)^T \xi(x_i)$ genau dann lokal c -optimal für einen schätzbaren Vektor $c \in \mathbb{R}^k$, wenn es eine Konstante γ und Konstanten $\varepsilon_i \in \{-1, 1\}$ gibt, so dass

gilt:

- $\gamma c = \sum_{i=1}^p \varepsilon_i \tilde{f}(x_i, \theta) \xi(x_i)$.
- γc ist ein Randpunkt von R .

Beweis

Spezialfall von Theorem 5.5 mit $s = 1$ oder Elfving (1952). \square

Effektiv verwendet ein c -optimales Design also genau die Trägerpunkte, mit deren Hilfe der Schnittpunkt des c -Vektors mit dem Rand des Elfving-Raums als Linearkombination konstruiert werden können. Die optimalen Gewichte sind genau die in dieser Linearkombination verwendeten Gewichte. Diese Repräsentation erlaubt es auch, eine Obergrenze für die Zahl der notwendigen Trägerpunkte zu formulieren:

Theorem 5.3

Zu jedem c -optimalen Design ξ für einen Vektor $c \in \mathbb{R}^k$ existiert ein Design $\tilde{\xi}$ mit höchstens k Trägerpunkten, so dass gilt:

$$-c^T M^-(\xi)c = -c^T M^-(\tilde{\xi})c.$$

Beweis:

Folgt aus Pázman (1986), Proposition III.16.

Es ist also ausreichend, die Suche nach c -optimalen Designs auf solche mit k oder weniger Trägerpunkten zu beschränken.

5.1.1 Beispiel im Modell mit fixen Effekten

Als erstes Beispiel betrachten wir wieder das nichtlineare Ein-Kompartiment-Modell der einfachen Evasion aus Abschnitt 4.2:

$$Y_i = \theta_1 e^{-\theta_2 x_i} + \epsilon_i, \quad i = 1, \dots, n$$

Für die Parameter gilt wie zuvor $\theta_1 = 46, \theta_2 = 1.7, \epsilon_i \sim \mathcal{N}(0, 1)$ unabhängig identisch, $c = (1, 10)$. Die Funktion \tilde{f} ist damit gegeben durch $\tilde{f}(x_i, \theta) = (e^{-\theta_2 x_i}, -x_i \theta_1 e^{-\theta_2 x_i})^T$. Der resultierende Elfving-Raum ist in Abbildung 5.1 dargestellt.

Nach Theorem 5.3 existiert ein c -optimales Design mit höchstens $k = 2$ Trägerpunkten. Aus Abbildung 5.1 erkennt man, dass zur Konvexkonstruktion des Schnittpunkts (blauer Kreis) des Randes des Elfving-Raums mit dem c -Vektor (rote Linie) gerade die zwei schwarz markierten Punkte mit den Funktionswerten $\tilde{f}(x_1, \theta) = (1, 0)^T$ und $\tilde{f}(x_2, \theta) = (-0.28, 9.7)^T$ notwendig sind. Entsprechend sind die optimalen Trägerpunkte diejenigen Werte x_1 und x_2 , für die gilt $\tilde{f}(x_1, \theta) = (1, 0)^T$ und $\tilde{f}(x_2, \theta) = (-0.28, 9.7)^T$. Dies sind gerade $x_1 = 0$ und $x_2 = 0.75$. Die Gewichte $\xi(x_1) = w$ und $\xi(x_2) = 1 - w$ können entweder graphisch bestimmt werden oder als Lösung der Gleichung $w\tilde{f}(x_1, \theta) + (1 - w)\tilde{f}(x_2, \theta) = \gamma c$ ermittelt werden (zwei Gleichungen für zwei Unbekannte w und γ). In diesem Fall erhält man $w = 0.554$. Das optimale Design ist daher das in Kapitel 4 betrachtete Design

$$\xi = \begin{pmatrix} 0 & 0.75 \\ 0.554 & 0.446 \end{pmatrix}.$$

5.2 Verallgemeinerter Satz von Elfving

Die klassische Version des Satzes von Elfving liefert eine geometrische Repräsentation optimaler Designs für Modelle mit Informationsmatrizen der Form

$$\mathcal{M}(\xi) = \sum_{i=1}^p \tilde{f}(x_i, \theta) \tilde{f}(x_i, \theta)^T \xi(x_i), \quad (5.2.1)$$

wobei ξ ein Design über \mathcal{X} mit p Trägerpunkten sei.

Nach Abschnitt 3.5 ist jedoch klar, dass diese Darstellung nicht für alle in dieser Arbeit betrachteten Modelle ausreichend ist. Sowohl für Modelle mit parameterabhängiger Varianz als auch für

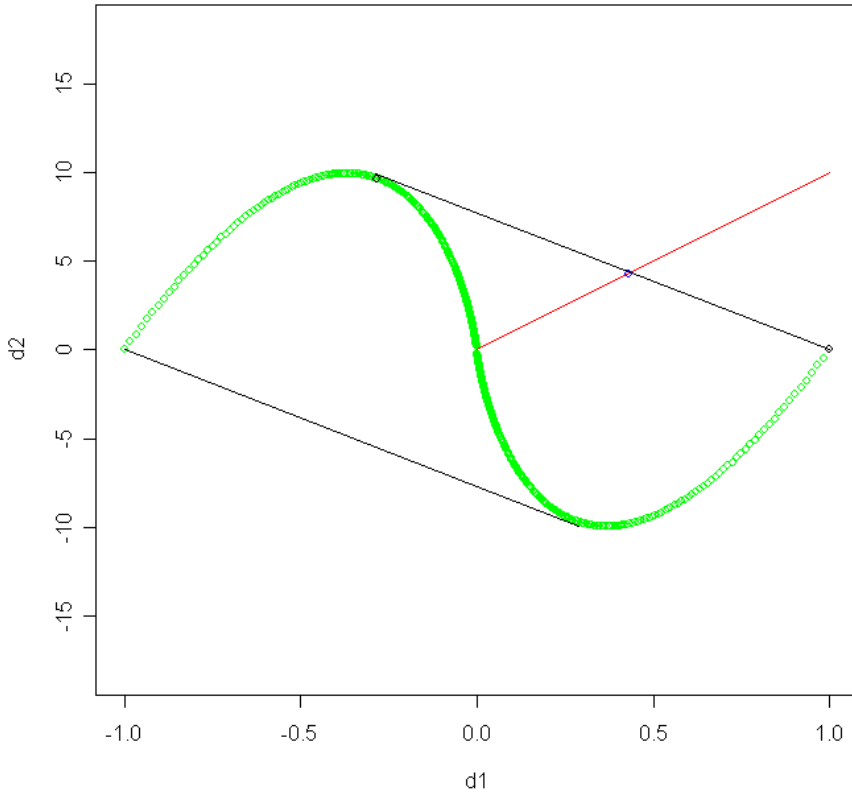


Abbildung 5.1: *Elfving Raum für ein homoskedastisches Ein-Kompartiment Modell mit fixen Effekten. Die grüne Kurve repräsentiert die durch den Versuchsbereich \mathcal{X} erzeugten Werte der Funktion \tilde{f} inklusive deren Spiegelung am Ursprung des Koordinatensystems. Der Elfving-Raum ist die konvexe Hülle dieser Kurve (markiert durch schwarze Linien). Die rote Linie repräsentiert den c -Vektor. Die schwarzen Punkte markieren die Vektoren $\tilde{f}(x_1, \theta)$ und $\tilde{f}(x_2, \theta)$, die zu den Trägerpunkten des optimalen Designs gehören. Der blaue Punkt markiert den Schnittpunkt des c -Vektors mit dem Rand des Elfving Raums.*

Modelle mit korrelierten Daten/Populationsmodelle liegt eine Informationsmatrix der Form

$$M(\xi) = \sum_{i=1}^p \sum_{l=1}^s \tilde{f}_l(x_i, \theta) \tilde{f}_l(x_i, \theta)^T \xi(x_i) \quad (5.2.2)$$

vor, mit $s = 2$ beim Maximum-Likelihood-Schätzer mit parameterabhängiger Varianz und $s = m$ beim NWLS-Schätzer im Populationsmodell mit m Beobachtungen pro Patient. Es ist

daher notwendig, den Satz von Elfving auf Modelle dieser Darstellung zu erweitern (vgl. Dette und Holland-Letz (2009)). Zu diesem Zweck definieren wir zunächst eine allgemeinere Version des Elfving Raums aus Definition 5.1:

Definition 5.4

Die Menge

$$\mathcal{R}_s = \text{conv} \left\{ \sum_{j=1}^s \varepsilon_j \tilde{f}_j(x, \theta) \mid x \in \mathcal{X}; \sum_{j=1}^s \varepsilon_j^2 = 1 \right\} \quad (5.2.3)$$

wird als verallgemeinerter Elfving-Raum bezeichnet.

Theorem 5.5

Bezeichne p_i die Wahrscheinlichkeitsmasse eines Designs ξ_c für den Punkt x_i , $i = 1, \dots, p$. Das Design $\xi_c = \{x_i, p_i\}_{i=1}^p$ ist in einem Modell mit Informationsmatrix der Form (5.2.2) genau dann lokal c -optimal für einen schätzbaren Vektor $c \in \mathbb{R}^k$, wenn Konstanten $\gamma > 0$, $\varepsilon_{11}, \dots, \varepsilon_{1p}, \dots, \varepsilon_{s1}, \dots, \varepsilon_{sp}$ existieren, so dass

$$\sum_{\ell=1}^s \varepsilon_{\ell r}^2 = 1; \quad r = 1, \dots, s \quad (5.2.4)$$

und so dass der Punkt $\gamma c \in \mathbb{R}^k$ im verallgemeinerten Elfving-Raums \mathcal{R}_s definiert in (5.2.3) ein Randpunkt ist, mit der Darstellung

$$\gamma c = \sum_{r=1}^p p_r \left\{ \sum_{\ell=1}^s \varepsilon_{\ell r} \tilde{f}_\ell(x_r, \theta) \right\}. \quad (5.2.5)$$

Beweis:

Angenommen, das Design $\xi_c = \{x_r; p_r\}_{r=1}^p$ minimiere $c^T M^{-1}(\xi) c$ oder äquivalent dazu erfülle Korollar 4.3. Wir definieren $\gamma^2 = (c^T G c)^{-1}$, $d = \gamma G c \in \mathbb{R}^k$, und es folgt aus $c \in \text{Range}(M(\xi))$ sowie der Darstellung (5.2.2) dass

$$\begin{aligned} \gamma c = M(\xi_c) d &= \sum_{r=1}^p p_r \left\{ \sum_{\ell=1}^s \tilde{f}_\ell(x_r, \theta) \tilde{f}_\ell^T(x_r, \theta) \right\} d \\ &= \sum_{r=1}^p p_r \left\{ \sum_{\ell=1}^s \tilde{f}_\ell(x_r, \theta) \varepsilon_{\ell r} \right\}, \end{aligned}$$

wobei die Notation $\varepsilon_{\ell r} = \tilde{f}_\ell^T(x_r, \theta)d$ ($\ell = 1, \dots, s; r = 1, \dots, p$) verwendet wurde. Nach Korollar 4.3 besteht Gleichheit in (4.1.2) an jedem Trägerpunkt x_r , woraus gefolgert werden kann (mittels Definition von γ und $\varepsilon_{\ell r}$), dass

$$\sum_{\ell=1}^s \varepsilon_{\ell r}^2 = 1; \quad r = 1, \dots, p.$$

Demnach sind die Bedingungen (5.2.4) und (5.2.5) erfüllt und es bleibt nur noch zu zeigen, dass γc ein Randpunkt von \mathcal{R}_s ist. Zu diesem Zweck nutzen wir aus, dass nach der Cauchy-Schwarz Ungleichung sowie Korollar 4.3 gilt, dass

$$\left(d^T \left\{ \sum_{\ell=1}^s \varepsilon_\ell \tilde{f}_\ell(x, \theta) \right\} \right)^2 \leq \sum_{\ell=1}^s (d^T \tilde{f}_\ell(x, \theta))^2 \sum_{\ell=1}^s \varepsilon_\ell^2 \leq \sum_{\ell=1}^s \varepsilon_\ell^2 = 1 \quad (5.2.6)$$

für alle $x \in \mathcal{X}$ und alle $\varepsilon_1, \dots, \varepsilon_\ell$ die (5.2.4) erfüllen. Darüber hinaus ist

$$\gamma c^T d = \gamma c^T \gamma G c = \gamma^2 c^T G c = (c^T G c)^{-1} c^T G c = 1$$

nach Definition der Konstanten γ und des Vectors d . Demnach definiert der Vektor d eine stützende Hyperebene für den verallgemeinerten Elfvingraum \mathcal{R}_s am Punkt γc , woraus folgt, dass γc ein Randpunkt von \mathcal{R}_s ist.

Zum Beweis der Rückrichtung nehmen wir an, dass γc ein Randpunkt ist und dass (5.2.4) und (5.2.5) erfüllt sind. In diesem Fall existiert eine stützende Hyperebene, bezeichnet als $d \in \mathbb{R}^k$, für den verallgemeinerten Elfving Raum \mathcal{R}_s an dem Randpunkt γc . Das bedeutet, es gilt

$$\gamma c^T d = 1, \quad (5.2.7)$$

und die Ungleichung

$$\left| d^T \left\{ \sum_{\ell=1}^s \varepsilon_\ell \tilde{f}_\ell(x, \theta) \right\} \right| \leq 1 \quad (5.2.8)$$

ist für alle $x \in \mathcal{X}$ und alle $\varepsilon_1, \dots, \varepsilon_s$ die (5.2.4) einhalten erfüllt. Definieren wir

$$\varepsilon_\ell(x) = \frac{d^T \tilde{f}_\ell(x, \theta)}{\sqrt{\sum_{j=1}^s (d^T \tilde{f}_j(x, \theta))^2}}, \quad \ell = 1, \dots, s$$

so gilt $\sum_{\ell=1}^s \varepsilon_\ell^2(x) = 1$ und aus der Ungleichung (5.2.8) ergibt sich

$$\sum_{\ell=1}^s (d^T \tilde{f}_\ell(x, \theta))^2 = \left| \sum_{\ell=1}^s d^T \{ \varepsilon_\ell(x) \tilde{f}_\ell(x, \theta) \} \right|^2 \leq 1 \quad (5.2.9)$$

für alle $x \in \mathcal{X}$. Andererseits folgt aus (5.2.5) und (5.2.7) dass

$$1 = \gamma c^T d = \sum_{r=1}^p p_r d^T \left\{ \sum_{\ell=1}^s \varepsilon_{\ell r} \tilde{f}_\ell(x_r, \theta) \right\},$$

was (mit Ungleichung (5.2.9)) wiederum für x_1, \dots, x_p bedingt, dass

$$d^T \left\{ \sum_{\ell=1}^s \varepsilon_{\ell r} \tilde{f}_\ell(x_r, \theta) \right\} = 1 \quad r = 1, \dots, p. \quad (5.2.10)$$

Demzufolge erhalten wir mittels der Cauchy-Schwarz-Ungleichung

$$1 = \left(\sum_{\ell=1}^s \varepsilon_{\ell r} d^T \tilde{f}_\ell(x_r, \theta) \right)^2 \leq \sum_{\ell=1}^s \varepsilon_{\ell r}^2 \sum_{\ell=1}^s \left(d^T \tilde{f}_\ell(x_r, \theta) \right)^2 \leq 1$$

für alle $r = 1, \dots, p$, und daraus wiederum

$$\varepsilon_{\ell r} = \lambda_r d^T \tilde{f}_\ell(x_r, \theta) \quad \ell = 1, \dots, s; r = 1, \dots, p \quad (5.2.11)$$

für Konstanten $\lambda_1, \dots, \lambda_p$. Einsetzen von (5.2.11) in (5.2.10) ergibt

$$1 = \lambda_r \sum_{\ell=1}^s \left(d^T \tilde{f}_\ell(x_r, \theta) \right)^2 = \lambda_r \quad r = 1, \dots, p,$$

und die Kombination von (5.2.11) und (5.2.5) resultiert in

$$\gamma c = \sum_{r=1}^p p_r \left\{ \sum_{\ell=1}^s \tilde{f}_\ell(x_r, \theta) \tilde{f}_\ell^T(x_r, \theta) \right\} d = M(\xi_c) d.$$

Aus Searle (1982) folgt die Existenz einer generalisierten Inverse der Matrix $M(\xi_c, \theta)$ so dass $d = \gamma G c$ und mit (5.2.9) gilt für alle $x \in \mathcal{X}$

$$\gamma^2 \sum_{\ell=1}^s (c^T G \tilde{f}_\ell(x))^2 \leq 1. \quad (5.2.12)$$

Nach (5.2.7) erhalten wir $\gamma^2 = (c^T G c)^{-1} = (c^T M^{-1}(\xi_c) c)^{-1}$ und die Ungleichung (5.2.12) vereinfacht sich zu (4.1.2). Die c -Optimalität des Designs $\xi_c = \{x_r; p_r\}_{r=1}^p$ folgt dann aus Korollar 4.3.

□

5.3 Beispiele

5.3.1 ML-Schätzer im Modell mit zufälligen Effekten

Als Beispiel für eine Informationsmatrix des Typs (5.2.2) betrachten wir einen Maximum-Likelihood-Schätzer im Modell mit zufälligen Effekten, d.h. das Modell

$$Y_i = \eta(x_i, b_i) + \varepsilon_i ; i = 1, \dots, n ,$$

mit unabhängig identisch normalverteilten Störgrößen mit Erwartungswert 0 und Varianz $\sigma^2 > 0$. Die unabhängigen Zufallsvariablen $b_1, \dots, b_n \sim \mathcal{N}(\theta, \Omega)$ bezeichnen jeweils den k -dimensionalen normalverteilten zufälligen Effekt. Wir nehmen an, dass auch b_1, \dots, b_n und $(\varepsilon_1, \dots, \varepsilon_n)^T$ stochastisch unabhängig sind. Die Varianz der Zufallsvariablen Y_i wird wie in (2.2.4) angenähert durch

$$\tilde{\sigma}^2(x_i, \theta) := \text{Var}(Y_i) \approx \tilde{f}(x_i, \theta)^T \Omega \tilde{f}(x_i, \theta) + \sigma^2. \quad (5.3.1)$$

und

$$\tilde{f}(x_i, \theta) = \frac{\partial \eta(x_i, \theta)}{\partial \theta}$$

Die Fisher Information für θ ist wie in (3.2.6) gegeben und hat dann die Struktur (5.2.2), d.h.

$$M(\xi) = \sum_{i=1}^p \sum_{l=1}^s \tilde{f}_l(x_i, \theta) \tilde{f}_l(x_i, \theta)^T \xi(x_i)$$

mit $s = 2$ und

$$\tilde{f}_1(x, \theta) = \frac{1}{\tilde{\sigma}(x, \theta)} \tilde{f}(x, \theta), \quad \tilde{f}_2(x, \theta) = \frac{1}{\sqrt{2\tilde{\sigma}^2(x, \theta)}} \frac{\partial \tilde{\sigma}^2(x, \theta)}{\partial \theta}.$$

Betrachten wir jetzt als Beispiel erneut das Einkompartiment-Modell mit intravenöser Verabreichung mit Konzentrationsfunktion

$$Y = b_1 e^{-b_2 x} + \varepsilon, \quad x \in \mathcal{X} = [0, 10], \quad (5.3.2)$$

und den Parametern

$$\theta = (30, 1.7) , \quad \Omega = \text{diag}(\omega_1, \omega_2) = \text{diag}(1, 0.1) \quad \text{und} \quad \sigma^2 = 0.04.$$

Direktes Ausrechnen liefert

$$\frac{\partial \eta(x, \theta)}{\partial \theta} = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T, \quad (5.3.3)$$

$$\frac{\partial^2 \eta(x, \theta)}{\partial \theta_1 \partial \theta_2} = -x e^{-\theta_2 x}, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_1^2} = 0, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_2^2} = \theta_1 x^2 e^{-\theta_2 x}, \quad (5.3.4)$$

und damit

$$\tilde{f}_1(x, \theta) = \frac{1}{\tilde{\sigma}(x, \theta)} (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T \quad (5.3.5)$$

$$\tilde{f}_2(x, \theta) = \frac{\sqrt{2}}{\tilde{\sigma}^2(x, \theta)} \sum_{m=1}^2 \left(\frac{\partial^2 \eta(x, \theta)}{\partial \theta_m \partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta_m} \right) \omega_m. \quad (5.3.6)$$

Der resultierende verallgemeinerte Elfving-Raum ist in Abbildung 5.2 dargestellt. Angenommen, es soll jetzt die Fläche unter der Kurve geschätzt werden, d.h.

$$AUC = \int_0^\infty \theta_1 e^{-\theta_2 x} dx = \frac{\theta_1}{\theta_2},$$

so erhalten wir das Problem des optimalen Designs für den Vektor $c = (1/\theta_2, -\theta_1/\theta_2^2)^T$. Der Vektor ist als rote Linie in Abbildung 5.2 dargestellt, sein Schnittpunkt mit dem Rand des Elfving-Raums als blauer Kreis und die notwendigen Trägerpunkte als schwarze Kreise. In dieser Situation ist die Identifikation der diesen Trägerpunkten zugrundeliegenden Versuchsbedingungen x analytisch nicht mehr möglich. Aus diesem Grund sind die Pixel entsprechend der zugehörigen Elemente $x \in \mathcal{X}$ eingefärbt. Die Zuordnung kann der Legende in Abbildung 5.2 entnommen werden, wobei die exakte Farbbestimmung anhand geeigneter Graphikprogramme erfolgen sollte.

Die Farbgebung der Trägerpunkte in Abbildung 5.2 ist lila und grün und entspricht den x -Werten 0.1 und 2.1. Wir erhalten als optimales Design für die AUC das Zwei-Punkt-Design

$$\xi_c = \begin{pmatrix} 0.1 & 2.1 \\ 0.24 & 0.76 \end{pmatrix}.$$

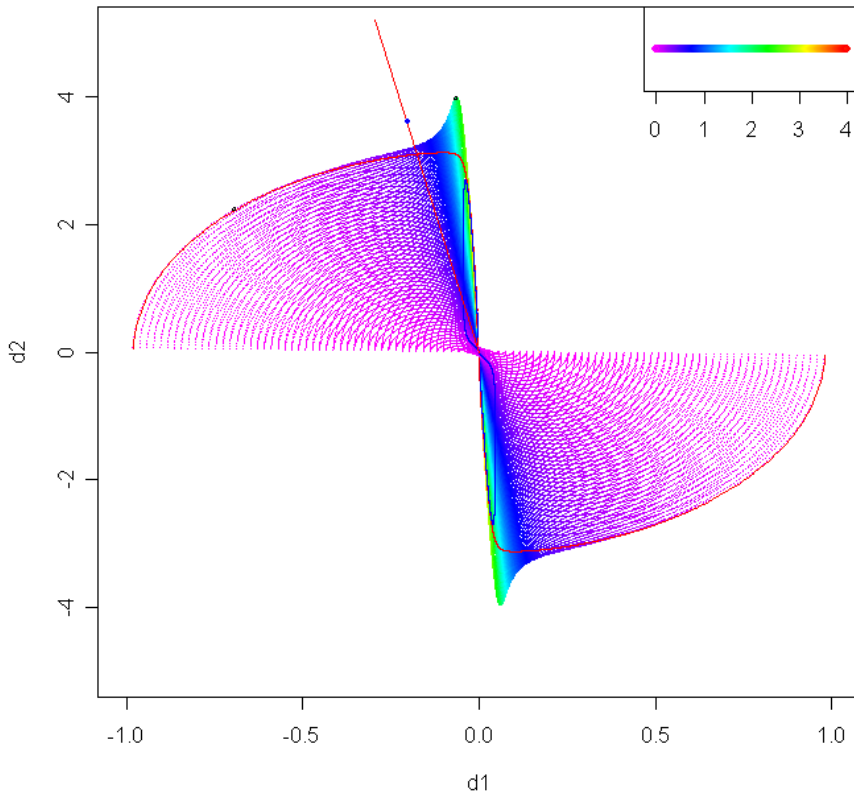


Abbildung 5.2: Der Elfving-Raum \mathcal{R}_2 definiert in (5.2.3) für den ML-Schätzers im Modell mit zufälligen Effekten für die einfache Evasionsfunktion. Die blaue und rote Kurve stellt die Funktionswerte der Funktionen \tilde{f}_1 und \tilde{f}_2 dar. Die daraus kombinierten Elemente des Elfving-Raums sind entsprechend den zugehörigen Versuchsbedingungen $x \in \mathcal{X}$ eingefärbt. Die Legende erlaubt die Zuordnung zu den Versuchsbedingungen. Die rote Linie beschreibt den c -Vektor. Der blaue Kreis markiert den Schnittpunkt des c -Vektors mit dem Rand des Elfving-Raums, schwarze Kreise stellen die notwendigen Trägerpunkte dar.

Als zweites Beispiel betrachten wir nun die Bateman Funktion, d.h. ein Modell mit drei Parametern

$$Y = b_1(e^{-b_2x} - e^{-b_3x}) + \varepsilon, \quad x \in \mathcal{X} = [0, 10], \quad (5.3.7)$$

und den Parametern

$$\theta = (46, 1.7, 1.2)^T, \quad \Omega = \text{diag}(\omega_1, \omega_2, \omega_3) = \text{diag}(1, 0.1, 0.1) \quad \text{und} \quad \sigma^2 = 0.04.$$

Die Ableitungen sind dann die folgenden:

$$\begin{aligned} \frac{\partial \eta(x, \theta)}{\partial \theta} &= (e^{-\theta_2 x} - e^{-\theta_3 x}, -\theta_1 x e^{-\theta_2 x}, \theta_1 x e^{-\theta_3 x})^T, \\ \frac{\partial^2 \eta(x, \theta)}{\partial \theta_1^2} &= 0, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_1 \partial \theta_2} = -x e^{-\theta_2 x}, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_1 \partial \theta_3} = x e^{-\theta_3 x}, \\ \frac{\partial^2 \eta(x, \theta)}{\partial \theta_2 \partial \theta_1} &= -x e^{-\theta_2 x}, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_2^2} = \theta_1 x^2 e^{-\theta_2 x}, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_2 \partial \theta_3} = 0, \\ \frac{\partial^2 \eta(x, \theta)}{\partial \theta_3 \partial \theta_1} &= x e^{-\theta_3 x}, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_3 \partial \theta_2} = 0, \quad \frac{\partial^2 \eta(x, \theta)}{\partial \theta_3^2} = -\theta_1 x^2 e^{-\theta_3 x}, \end{aligned}$$

und damit

$$\tilde{f}_1(x, \theta) = \frac{1}{\tilde{\sigma}(x, \theta)} (e^{-\theta_2 x} - e^{-\theta_3 x}, -\theta_1 x e^{-\theta_2 x}, \theta_1 x e^{-\theta_3 x})^T \quad (5.3.8)$$

$$\tilde{f}_2(x, \theta) = \frac{\sqrt{2}}{\tilde{\sigma}^2(x, \theta)} \sum_{m=1}^3 \left(\frac{\partial^2 \eta(x, \theta)}{\partial \theta_m \partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta_m} \right) \omega_m. \quad (5.3.9)$$

Der resultierende verallgemeinerte Elfving-Raum ist in Abbildung 5.3 dargestellt. Angenommen, es soll wieder die Fläche unter der Kurve geschätzt werden, d.h.

$$AUC = \int_0^\infty \theta_1 (e^{-\theta_2 x} - e^{-\theta_3 x}) dx = \frac{\theta_1}{\theta_2} - \frac{\theta_1}{\theta_3},$$

so erhalten wir das Problem des optimalen Design für den Vektor $c = (\frac{1}{\theta_2} - \frac{1}{\theta_3}, -\frac{\theta_1}{\theta_2}, \frac{\theta_1}{\theta_3})^T$. Der Vektor ist als rote Linie in Abbildung 5.3 dargestellt.

Hier das optimale Design zu finden erfordert eine Darstellung des Elfving-Raums als rotierende 3D-Graphik (nicht dargestellt). Man erhält als optimales Design das Drei-Punkt-Design

$$\xi_c = \begin{pmatrix} 0.1 & 0.9 & 2.9 \\ 0.77 & 0.19 & 0.04 \end{pmatrix}.$$

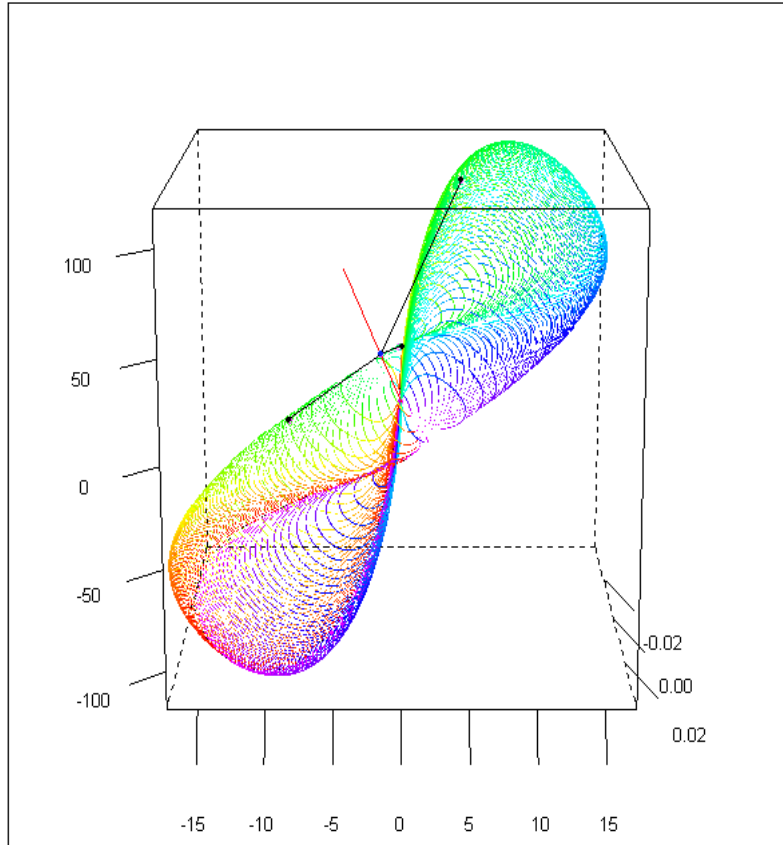


Abbildung 5.3: Der Elfving-Raum \mathcal{R}_2 definiert in (5.2.3) für den ML-Schätzers im Modell mit zufälligen Effekten für die Bateman Funktion. Die rote Linie beschreibt den c -Vektor. Der blaue Kreis markiert den Schnittpunkt des c -Vektors mit dem Rand des Elfving-Raums, schwarze Kreise stellen die notwendigen Trägerpunkte dar.

5.3.2 NWLS-Schätzer im linearen Modell mit korrelierten Beobachtungen

Die zweite behandelte Situation mit Informationsmatrizen der Form (5.2.2) war der NWLS Schätzer im Modell mit m korrelierten Beobachtungen an einer Anzahl N verschiedener Patienten (vgl. Holland-Letz et al. (2009)). Wir beginnen hier mit einem quadratischen linearen Modell mit fixen Effekten und korrelierten Störgrößen:

$$Y_{ij} = \theta_1 x_{ij} + \theta_2 x_{ij}^2 + \epsilon_{ij} \quad i = 1, \dots, N, j = 1, \dots, m. \quad (5.3.10)$$

Wir setzen zunächst $m = 2$ und nehmen weiterhin an, dass Störgrößen an demselben Patienten korreliert sind mit Kovarianzfunktion $\text{cov}(\epsilon_{i1}, \epsilon_{i2}) := \sigma^2 \text{co}(x_{i1}, x_{i2}) = \sigma^2 \lambda^{|x_{i1} - x_{i2}|}$, $\lambda \in [0, 1]$, d.h.

$$V_i^{-1} = \sigma^{-2} \begin{pmatrix} 1 & \lambda^{|x_{i1} - x_{i2}|} \\ \lambda^{|x_{i1} - x_{i2}|} & 1 \end{pmatrix}^{-1} = \begin{pmatrix} v_{11}(x_i) & v_{12}(x_i) \\ v_{21}(x_i) & v_{22}(x_i) \end{pmatrix}.$$

In diesem Modell gilt $f(x) = (x, x^2)^T$, $F_i = (f(x_{i1}), f(x_{i2}))$ und die notwendigen Funktionen $\tilde{f}_\ell(x)$, $\ell = 1, 2$ sind gegeben als Spalten der Matrix

$$F_i^* = \tilde{F}_i V_i^{-0.5}. \quad (5.3.11)$$

Seien jetzt als Parameter $\lambda = 0.6$, $\sigma^2 = 0.04$, $\theta = (\theta_1, \theta_2)^T = (3, 1.7)^T$ und der Versuchsbereich $\mathcal{X} = [0, 2]$ vorgegeben, so ist der verallgemeinerte Elfving-Raum \mathcal{R}_2 definiert in (5.2.3) wie in Abbildung 5.4 dargestellt. Jeder Pixel dieser Darstellung wird erzeugt durch einen spezifischen Satz von m Messungen unter Versuchsbedingungen $x \in \mathcal{X}^m$ ($m = 2$), wobei die Funktionen \tilde{f}_ℓ und die Größen ϵ_ℓ in (5.2.3) entlang eines engmaschigen Gitters über den Versuchsbereich \mathcal{X}^m bzw. das Intervall $[-1, 1]$ abgetastet werden.

Beide Teile der Abbildung stellen denselben Elfving-Raum dar, allerdings korrespondiert die Farbgebung der linken Seite zu den potentiellen Werten der ersten Komponente x_1 eines Messsatzes $x = (x_1, x_2)$, während die Farbgebung der rechten Seite der zweiten Komponente x_2 entspricht (Siehe Legende in Abbildung 5.4).

Angenommen, es soll jetzt die Linearkombination $c^T \theta$ definiert durch den Vektor $c = (-1, 1)^T$ geschätzt werden (eingetragen als rote Linie in Abbildung 5.4). Die optimalen Messsätze sind diejenigen, mit denen der exakte Schnittpunkt des c -Vektors mit dem Rand des Elfving-Raums als Konvexkombination konstruiert werden kann. Diese Konstruktion erfordert entweder einen einzelnen Punkt der Art

$$p(x) = \sum_{\ell=1}^m \epsilon_\ell \tilde{f}_\ell(x); \quad \sum_{\ell=1}^m \epsilon_\ell^2 = 1$$

oder mehrere Punkte $p(x_1), \dots, p(x_p)$ desselben Types. Im Höchstfall sind $p = k$ Punkte notwendig, wobei k die Zahl der Parameter des Modells darstellt (vgl. Theorem 5.3). In diesem Modell ist $k = 2$. Jeder Punkt $x_j = (x_{j1}, \dots, x_{jm}) \in \mathcal{X}^m$ entspricht dabei einem Satz von $m = 2$ Messungen an einem individuellen Patient. Die in der Konvexkombination verwendeten Gewichte ergeben

dann die entsprechenden Gewichte im optimalen Design, das heißt die relativen Häufigkeiten der Patienten, die unter dem entsprechenden Satz Messbedingungen $x_j = (x_{j1}, x_{j2})$ beobachtet werden sollten. Die zugrundeliegenden Komponenten x_{j1} und x_{j2} können anhand der Farbgebung der Punkte $p(x_j)$ im linken bzw. im rechten Teil von Abbildung 5.4 bestimmt werden. Die Trägerpunkte und Gewichte des optimalen Designs können demnach graphisch ermittelt werden. Betrachten wir jetzt als Beispiel Abbildung 5.4, so erkennt man, dass 2 Punkte x_1 und x_2 notwendig sind um den Randpunkt γc darzustellen. Diese Punkte sind durch zwei schwarze Kreise markiert. Am linken Teil der Graphik erkennen wir, dass die Farben der Punkte x_1 und x_2 gerade pink und grün sind. Mit Hilfe der Legende in der oberen Ecke können die zugrundeliegenden Werte identifiziert werden, und wir erhalten $x_{11} = 0.0$ und $x_{21} = 1.2$ als erste Komponenten von x_1 und x_2 . Auf dieselbe Weise ergibt die rechte Seite von Abbildung 5.4 die Farben blau und rot, d.h. die zweiten Komponenten sind $x_{12} = 0.8$ und $x_{22} = 2.0$.

Zusammenfassend weist das lokal c -optimale Design den Untersucher daher an, zwei verschiedene Messsätze zu verwenden, nämlich:

$$\xi_c = \begin{pmatrix} (0.0, 0.8) & (1.2, 2.00) \\ 0.48 & 0.52 \end{pmatrix}. \quad (5.3.12)$$

Es werden also 48% der Patienten unter Versuchsbedingungen $x_{11} = 0, x_{12} = 0.8$ behandelt, und 52% unter $x_{21} = 1.2$ und $x_{22} = 2.0$. Anzumerken bleibt, dass in der praktischen Anwendung mittels geeigneter Graphiksoftware die genaue Farbe der fraglichen Punkte präzise anhand der rot/grün/blau Werte der Pixel bestimmt werden kann.

Als Alternative zur direkten graphischen Punktbestimmung ist es ebenfalls möglich, anhand der Graphik lediglich eine stützende Hyperbene H des Elfving-Raums zu identifizieren, mit dem stützenden Randpunkt γc . Eine solche Ebene ist definiert durch einen Vektor $d = (d_1, d_2)^T$ und erfüllt $d^T z = 1$ für alle $z \in H$, $(\gamma c)^T d = 1$ und $r^T d \leq 1$ für alle $r \in \mathcal{R}_2$. Die Trägerpunkte des optimalen Designs sind dann gegeben durch die Lösungen des Gleichungssystems

$$\max_{\epsilon_1, \epsilon_2} \sum_{l=1}^2 \epsilon_l \tilde{f}_l(x_1, x_2)^T d = 1, \quad \sum_{i=1}^2 \epsilon_i^2 = 1 \quad (5.3.13)$$

[Vergleiche Gleichung (5.2.6) im Beweis von Theorem 5.5]. Im Beispiel ist der Vektor d gegeben durch $d = (-0.31, 0.16)^T$. Tastet man jetzt den Versuchsbereich \mathcal{X}^2 durch einen zweidimensio-

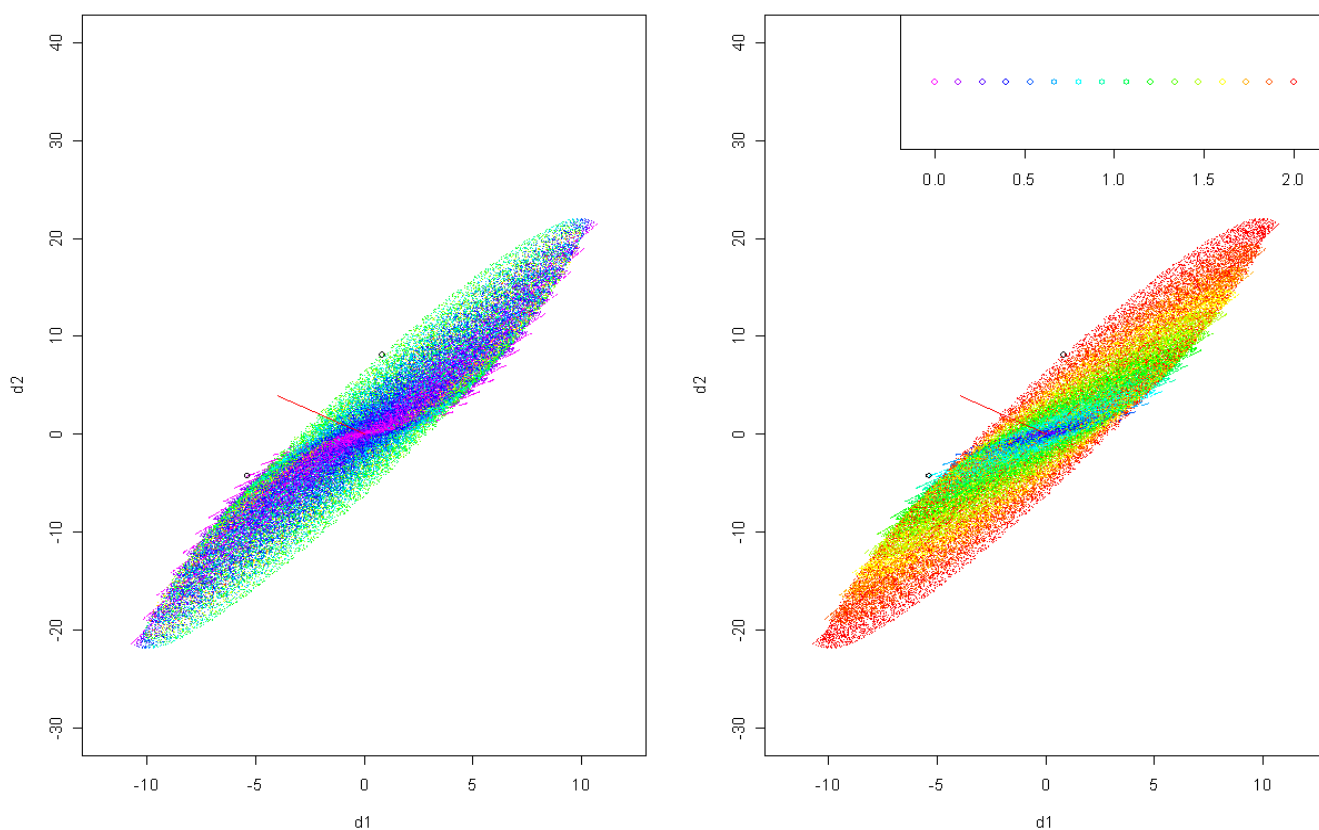


Abbildung 5.4: Der Elfvig-Raum \mathcal{R}_2 definiert in (5.2.3) für ein quadratisches Regressionsmodell mit zwei Beobachtungen pro Patient. Die Funktionen \tilde{f}_1 und \tilde{f}_2 zur Konstruktion der Einträge sind gegeben durch die Spalten von (5.3.11). Der Vektor c ist durch eine rote Linie dargestellt, die in der Elfvig-Repräsentation (5.2.5) verwendeten Punkte durch zwei schwarze Kreise. Die Farbgebung in der linken bzw. der rechten Abbildung entspricht dem ersten bzw. zweiten Eintrag des zugrundeliegenden Sets an Versuchsbedingungen.

nalen Vektor $u = [(0, 0.1)^T, (0, 0.2)^T, \dots, (0, 2.0)^T, (0.1, 0.2)^T, \dots, (0.1, 2.0)^T, \dots, (1.9, 2.0)^T]$ ab, so kann dieser gegen das Ergebnis der linken Seite von Gleichung (5.3.13) dargestellt werden (siehe Abbildung 5.5).

Man erkennt, dass der Wert in Gleichung (5.3.13) stets kleiner/gleich 1 ist, mit Gleichheit für die Einträge 8 und 182. Diese entsprechen genau den Werten $x_1 = (0, 0.8)$ und $x_2 = (1.2, 2.0)$.

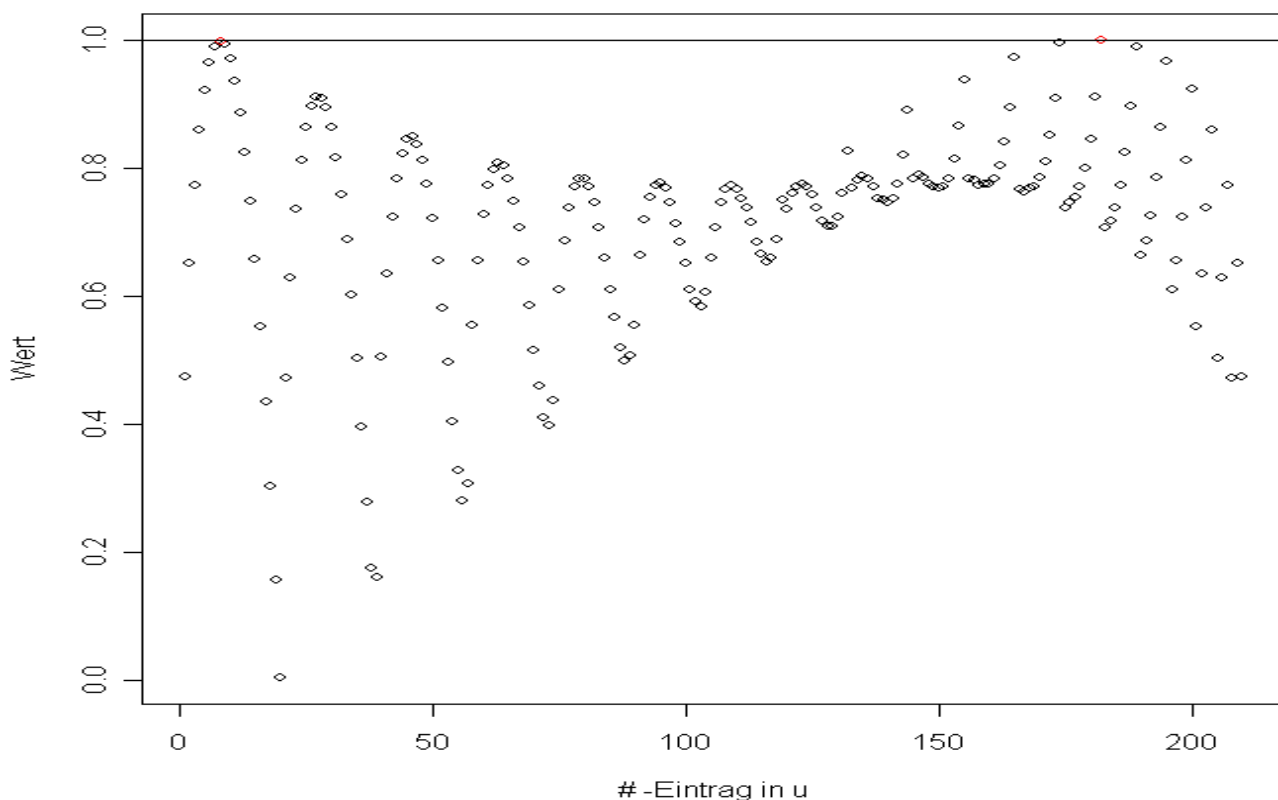


Abbildung 5.5: *Bestimmung der Trägerpunkte des optimalen Designs anhand der Hyperebenengleichung. Einträge von u mit Wert 1 (rot markiert) entsprechen den optimalen Trägerpunkten*

Auf diese Weise ist also eine alternative Bestimmung des Designs (5.3.12) möglich.

In beiden Fällen kann die Optimalität des Designs mittels Korollar 4.3 bestätigt werden.

Wir nehmen jetzt an, dass im selben Modell $m = 3$ Beobachtungen zur Verfügung stehen. Es gilt weiterhin $f(x) = (x, x^2)^T$, die Designmatrix ist jetzt $F_i = (f(x_{i1}), f(x_{i2}), f(x_{i3}))$ und die Funktionen $\tilde{f}_\ell(x)$, $\ell = 1, 2, 3$ sind die Spalten von

$$F_i^* = \tilde{F}_i V_i^{-0.5}. \quad (5.3.14)$$

Die Matrix $V_i^{-1} = (v_{lk}(x_i))_{l,k=1}^3$ ist gegeben durch

$$V_i^{-1} = \sigma^{-2} \begin{pmatrix} 1 & \lambda^{|x_{i1}-x_{i2}|} & \lambda^{|x_{i1}-x_{i3}|} \\ \lambda^{|x_{i2}-x_{i1}|} & 1 & \lambda^{|x_{i2}-x_{i3}|} \\ \lambda^{|x_{i3}-x_{i1}|} & \lambda^{|x_{i3}-x_{i2}|} & 1 \end{pmatrix}^{-1}.$$

Der entstehende Elfving-Raum ist in Abbildung 5.6 dargestellt. Da jetzt $m = 3$ ist, sind drei Teilabbildungen notwendig, zugehörig zu den drei Komponenten $x_i = (x_{i1}, x_{i2}, x_{i3})$. Unter diesen Vorgaben wird nur ein Punkt in der Elfving-Repräsentation (5.2.5) verwendet, und mit identischem Vorgehen wie zuvor erhält man für den Vektor $c = (-1, 1)^T$ das c -optimale Design

$$\xi_c = \begin{pmatrix} (0.0, 1.0, 2.0) \\ 1 \end{pmatrix}.$$

Alle Patienten sind demnach einheitlich unter den Versuchsbedingungen 0, 1.0 und 2.0 zu behandeln.

5.3.3 NWLS-Schätzer im Populationsmodell

Als letztes Beispiel betrachten wir jetzt wiederum die einfache Evasion im Populationsmodell, d.h. ($b_i = (b_{i1}, b_{i2})$)

$$Y_{ij} = b_{i1}e^{-b_{i2}x_{ij}} + \varepsilon_{ij}, \quad x_{ij} \in \mathcal{X} = [0, 2], i = 1, \dots, N, j = 1, \dots, m, \quad (5.3.15)$$

Die Störgrößen seien hier als homoskedastisch und unkorreliert angenommen, das heißt für die Varianz der Beobachtungen $Y_i = (Y_{i1}, \dots, Y_{im})$ am Patienten i gilt

$$\text{Var}(Y_i) = V_i \approx \tilde{F}_i^T \Omega \tilde{F}_i + \sigma^2 I_m \quad i = 1, \dots, N.$$

Als Parameter nehmen wir diesmal an

$$\theta = (5, 0.08)^T, \quad \Omega = \text{diag}(1, 0.1) \text{ and } \sigma^2 = 0.04,$$

und es stehen $m = 2$ Beobachtungen pro Patient zur Verfügung.

Für die Ableitungen der Konzentrationsfunktion gilt

$$\frac{\partial \eta(x, \theta)}{\partial \theta} = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T.$$

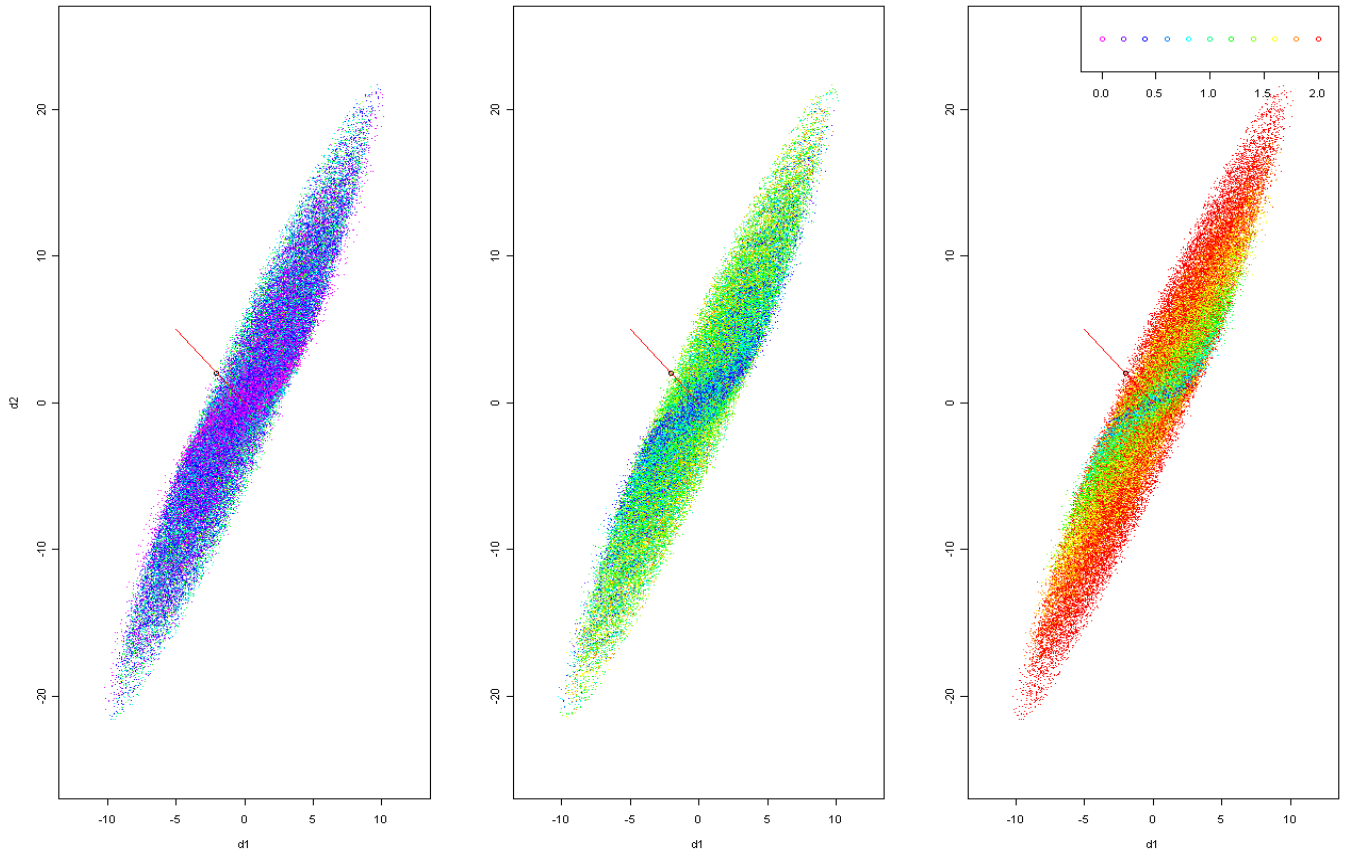


Abbildung 5.6: Der Elfving-Raum \mathcal{R}_3 definiert in (5.2.3) für ein quadratisches Regressionsmodell mit drei Beobachtungen pro Patient. Der Vektor c ist durch eine rote Linie dargestellt, der in der Elfving-Repräsentation (5.2.5) verwendete Punkt durch einen schwarzen Kreis.

und wir erhalten $f(x, \theta) = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})$. Die zwei Funktionen \tilde{f}_1, \tilde{f}_2 sind dann wieder die Spalten der in (5.3.11) definierten Matrix F^* , unter Verwendung der veränderten Funktion f . Der resultierende Elfving-Raum ist in Abbildung 5.7 dargestellt. Die Schätzung der Fläche unter der Konzentrationsfunktion entspricht auch hier der Suche nach einem c -optimalen Design für den c -Vektor $c = (1/\theta_2, -\theta_1/\theta_2^2)^T$, der als rote Linie in Abbildung 5.7 eingetragen ist. Es wird deutlich, dass nur ein einzelner Messsatz zur Schätzung notwendig ist, und man erhält als c -optimales Design zur Schätzung der Fläche unter der Kurve das Design

$$\xi_c = \begin{pmatrix} (0.0, 2.0) \\ 1 \end{pmatrix}.$$

Alle Patienten sollten also unter Versuchsbedingungen $x_1 = 0$ und $x_2 = 2$ behandelt werden. Die Optimalität dieses Designs kann mit Korollar 4.3 bestätigt werden.

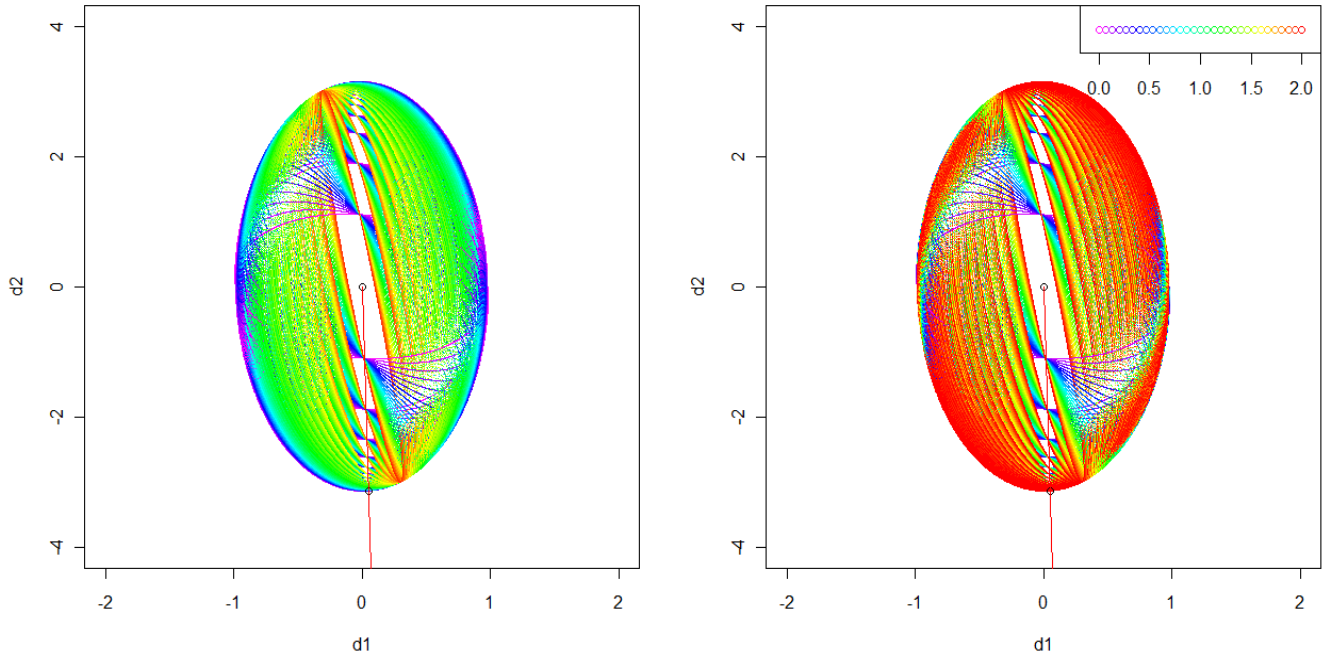


Abbildung 5.7: Der Elfving-Raum \mathcal{R}_2 definiert in (5.2.3) für das Einkompartiment-Modell mit intravenöser Verabreichung mit zwei Beobachtungen pro Patient. Der Vektor c ist durch eine rote Linie dargestellt, der in der Elfving-Repräsentation (5.2.5) verwendete Punkt durch einen schwarzen Kreis.

Kapitel 6

Algorithmen

Optimale Designs für konkrete Probleme lassen sich nur in den seltensten Fällen direkt analytisch bestimmen. In vielen Situationen erlauben es graphische Repräsentationen (Kapitel 5) die Designs zu ermitteln, aber speziell bei hochdimensionaleren Modellen stoßen diese Methoden schnell an ihre Grenzen. Für sehr allgemeine Situationen erlauben es Äquivalenzsätze, zumindest die Optimalität eines potentiellen Designs zu bewerten. Ausgehend von diesen Erkenntnissen ist es möglich, Algorithmen zu formulieren, die anhand eines über den Versuchsbereich gelegten Rasters iterativ (annähernd) optimale Designs bestimmen können.

6.1 Ein multiplikativer Algorithmus für c-optimale Designs

Gegeben ein endlicher Versuchsbereich $\mathcal{X} = \{x_1, \dots, x_r\}$ und Wahrscheinlichkeitsmaße w über \mathcal{X} mit $w(x_i) = w_i$. Bezeichne $S = \{w : w_i > 0, \sum_{i=1}^r w_i = 1\}$, und sei w^0 ein Element aus S als Startwert des Algorithmus. Weiterhin bezeichne $M(w) = M$ die zugehörige Informationsmatrix, A_i die Informationsmatrix einer einzelnen Messung am Punkt x_i , $i = 1, \dots, r$ und $\phi(M)$ eine zu minimierende Informationsfunktion. $M(w)$ sei als invertierbar angenommen für alle $w \in S$.

Dann definiere

$$w_i^{q+1} = w_i^q \frac{d_i^\lambda}{\sum_{j=1}^r w_j^q d_j^\lambda} \quad (6.1.1)$$

mit $\lambda \in (0, 1]$ und

$$d_i = \text{tr}\left(\frac{\partial\phi(M)}{\partial M} A_i\right) \quad (6.1.2)$$

den $q + 1$ - Schritt des Algorithmus.

Für einen allgemeinen Beweis der Monotonie dieses Algorithmus in ϕ für alle λ , siehe Yu (2009), ebenso wie für Konvergenz im Falle $\lambda < 1$. In jedem Schritt liefert der Algorithmus also mindestens gleich gute Designs wie zuvor, zumindest für $\lambda < 1$ wird auch das optimale Design beliebig genau angenähert. Für den Fall $\lambda = 1$ ist die Konvergenz nur für bestimmte Informationsfunktionen allgemein gegeben (z.B. D-Optimalität, siehe Pázman (1986) oder Dette et al. (2008b)). Im Falle von c -Optimalität existieren Gegenbeispiele (z.B. Beispiel 6.1.1). Die Verwendung von $\lambda < 1$ ist jedoch immer möglich.

Lemma 6.1

Für ϕ^p -Optimalität gilt

$$d_i = \text{tr}(M^{-1}K(K^T M^{-1}K)^{-p-1}K^T M^{-1}A_i).$$

und damit

$$w_i^{q+1} = w_i^q \frac{\text{tr}(M^{-1}K(K^T M^{-1}K)^{-p-1}K^T M^{-1}A_i)^\lambda}{\sum_{j=1}^r w_j^q \text{tr}(M^{-1}K(K^T M^{-1}K)^{-p-1}K^T M^{-1}A_j)^\lambda}. \quad (6.1.3)$$

Beweis

Ein ϕ^p -optimales Design minimiert $\text{tr}((K^T M^{-1}K)^{-p})$. Verwendet man die Definition des Differentials $d\phi(X)$ durch $d\phi(X) = \frac{\partial\phi(X)}{\partial \text{vec } X} \text{vec } dX$ (siehe Magnus und Neudecker (1999), $\frac{\partial\phi(X)}{\partial \text{vec } X}$ ist ein Zeilenvektor) und die Beziehungen

1. $d\text{tr}(X) = \text{tr}(dX)$

2. $dtr(X^p) = tr(p(X^T)^{p-1}dX)$
3. $dAXB = A(dX)B$
4. $dX^{-1} = -X^{-1}(dX)X^{-1}$
5. $d\phi(X) = tr(AdX) \Rightarrow \frac{\partial\phi(X)}{\partial X} = A^T$

(siehe Kapitel 9, Magnus und Neudecker (1999)), erhält man:

$$\begin{aligned}
dtr[(K^T M^{-1} K)^{-p}] &= tr[d(K^T M^{-1} K)^{-p}] \\
&= tr[-p(K^T M^{-1} K)^{-p-1}d(K^T M^{-1} K)] \\
&= tr[-p(K^T M^{-1} K)^{-p-1}K^T(dM^{-1})K] \\
&= tr[p(K^T M^{-1} K)^{-p-1}K^T M^{-1}(dM)KM^{-1}] \\
&= tr[pKM^{-1}(K^T M^{-1} K)^{-p-1}K^T M^{-1}(dM)]
\end{aligned}$$

und mit Beziehung 5. gilt also

$$\frac{\partial\phi^p(M)}{\partial M} = p M^{-1}K(K^T M^{-1}K)^{-p-1}K^T M^{-1}. \quad (6.1.4)$$

Einsetzen in die Darstellung (6.1.2) liefert die Behauptung. \square

Korollar 6.2

Für einen Vektor $c \in \mathbb{R}^k$, das Kriterium der c -Optimalität und $\lambda = 1$ gilt:

$$w_i^{q+1} = w_i^q \frac{c^T M^{-1} A_i M^{-1} c}{c^T M^{-1} c}. \quad (6.1.5)$$

Für Informationsmatrizen der Form $A_i = \sum_{\ell=1}^m \tilde{f}_\ell(x_i) \tilde{f}_\ell(x_i)^T$ gilt:

$$w_i^{q+1} = w_i^q \sum_{\ell=1}^m \frac{(\tilde{f}_\ell(x_i) M^{-1} c)^2}{c^T M^{-1} c}.$$

Beweis

Folgt direkt aus Lemma 6.1 mit $\lambda = 1$, $p = 0$ und $K = c$, mit der Beziehung

$$\begin{aligned}
&tr[\sum_{j=1}^r w_j^q M^{-1} K (K^T M^{-1} K)^{-p-1} K^T M^{-1} A_j] \\
&= tr[MM^{-1}K(K^T M^{-1}K)^{-1}K^T M^{-1}] = 1 \quad \square.
\end{aligned}$$

Der Multiplikator der Gewichte entspricht also bei $\lambda = 1$ der linken Seite der Normalenungleichung in Korollar 4.2 und 4.3.

Die Anwendung des Algorithmus ist in allen in dieser Arbeit behandelten Situationen mit existierender Darstellung approximativer Designs problemlos möglich. Nicht anwendbar ist er dagegen bei Problemstellungen ohne solche Darstellungen.

6.1.1 Beispiel: Unkorrelierte Beobachtungen

Als erstes Beispiel betrachten wir erneut die Situation des gewichteten Kleinste-Quadrate-Schätzers im Einkompartimentmodell mit intravenöser Verabreichung aus Beispiel 4.2. Wie zuvor liegt das Modell in der Form (2.1.3) vor:

$$Y_i = \theta_1 e^{-\theta_2 x_i} + \epsilon_i, \quad i = 1, \dots, n \quad (6.1.6)$$

Als Parameter seien wiederum $\theta_1 = 46, \theta_2 = 1.7, \epsilon_i \sim \mathcal{N}(0, 1)$ unabhängig identisch, $i = 1, \dots, n$ sowie $c = (1, 10)$ vorgegeben, und $\tilde{f}(x_i, \theta) = (e^{-\theta_2 x_i}, -x_i \theta_1 e^{-\theta_2 x_i})^T$. Der Versuchsbereich \mathcal{X} sei als $\mathcal{X} = (0, 0.01, 0.02, \dots, 2.99)$ festgelegt, d.h. $r = 300$

Wir beginnen mit einem gleichverteilten Ausgangsdesign w^0 mit $w_i^0 = \frac{1}{300}$ für alle $i = 1, \dots, 300$ und führen Iterationen des Algorithmus 6.1.1 mit $\lambda = 0.99$ durch. Die erzielten Gewichte nach 10 bzw. 1000 Iterationen sind in Abbildung 6.1 dargestellt.

Offensichtlich sieht der Algorithmus vor, 55.4% der Beobachtungen zum Zeitpunkt 0 durchzuführen, während sich die restlichen 44.6% in der Umgebung des Punktes 0.75 verteilen. Das Design kann so verwendet werden oder durch das bereits bekannte Zweipunkt-Design

$$\xi = \begin{pmatrix} 0 & 0.75 \\ 0.554 & 0.446 \end{pmatrix}.$$

angenähert werden. Mittels Korollar 4.2 kann geprüft werden, ob dieses Design tatsächlich optimal ist. In Abbildung 6.2 ist die linke Seite von 4.1.2 für alle $x \in \mathcal{X}$ dargestellt, und man erkennt, dass hier in der Tat die Bedingungen für ein optimales Design erfüllt werden.

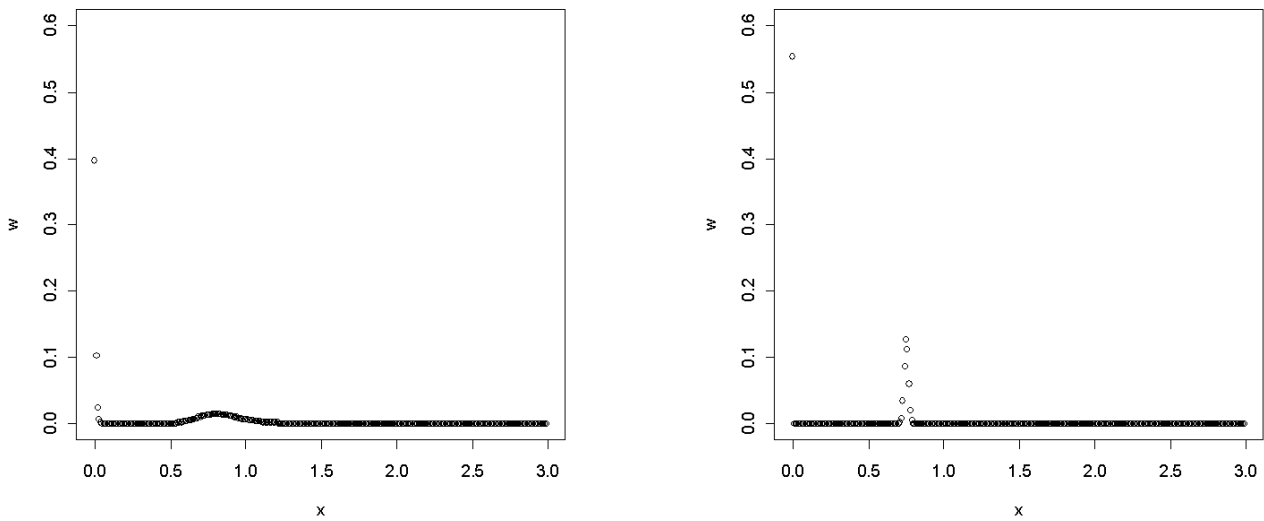


Abbildung 6.1: Verteilung der Gewichte nach 10 (links) bzw. 100 (rechts) Durchläufen des Algorithmus 6.1.1 im Modell 6.1.6.

6.1.2 Beispiel: Korrelierte Beobachtungen

Als zweites Beispiel betrachten wir ebenfalls für den gewichteten kleinste-Quadrate-Schätzer das Modell einer Bateman-Funktion (2.4.5) im Populationsmodell mit zwei zulässigen Beobachtungen pro Patient. Das Modell ist das folgende:

$$Y_{ij} = b_1(e^{-b_2x_{ij}} - e^{-b_3x_{ij}}) + \epsilon_{ij}, \quad i = 1, \dots, N, j = 1, \dots, m \quad (6.1.7)$$

Als Parameter seien $\theta = (\theta_1, \theta_2, \theta_3) = (46, 1.7, 1.2)$ gegeben. Wir nehmen an, dass wie in Beispiel 5.3.2 die Beobachtungen $Y_i = (Y_{i1}, \dots, Y_{im})^T$ am selben Patienten korreliert sind mit Kovarianzfunktion

$$\text{Var}(Y_i) = V_i \approx \tilde{F}_i^T \Omega \tilde{F}_i + \sigma^2 I_m \quad i = 1, \dots, N.$$

mit $\sigma > 0$, für dieses Beispiel gewählt als $\sigma^2 = 0.04$ und $\Omega = \text{diag}(1, 0.1, 0.1)$. Auch hier bezeichne $v_{j,l}$ ($j, l = 1, \dots, m$) die Elemente der Inversen der Kovarianzmatrix V_i .

Geschätzt werden soll in diesem Fall die Fläche unter der Konzentrationskurve. Nach 3.5.7 ist daher

$$c = \left(\frac{1}{\theta_2} - \frac{1}{\theta_3}, -\frac{\theta_1}{\theta_2^2}, \frac{\theta_1}{\theta_3^2} \right).$$

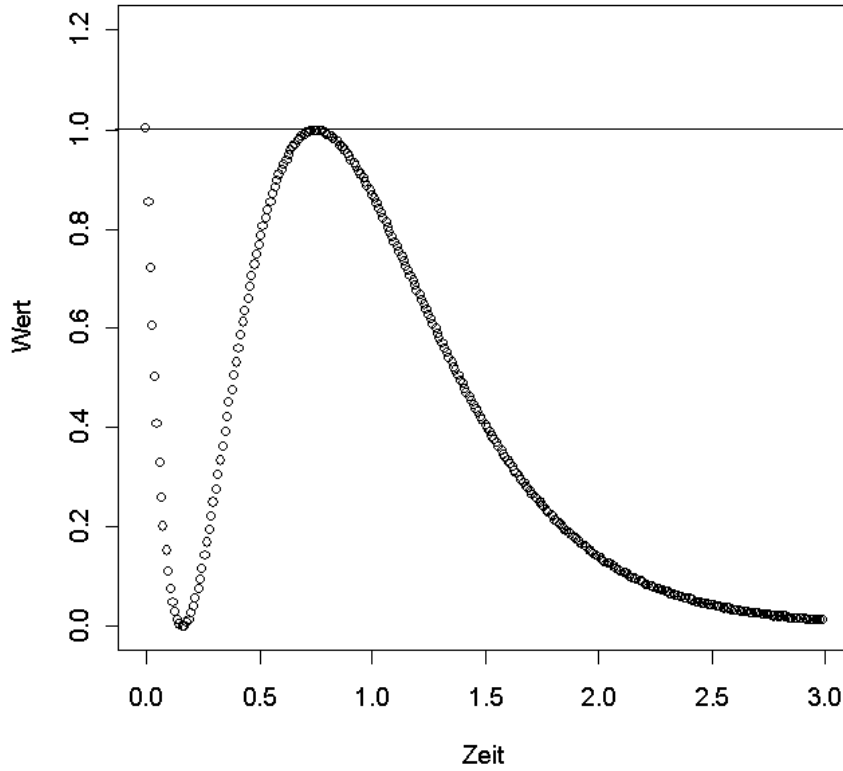


Abbildung 6.2: Darstellung des Äquivalenzsatzes 4.1.2 für das mittels Algorithmus 6.1.1 im Modell 6.1.6 erzeugte Design.

Die Ableitungen der Konzentrationsfunktion sind gegeben durch

$$\frac{\partial \eta(x, \theta)}{\partial \theta} = \tilde{f}(x, \theta) = (e^{-\theta_2 x} - e^{-\theta_3 x}, -\theta_1 x e^{-\theta_2 x}, \theta_1 x e^{-\theta_3 x})^T,$$

und die Designmatrix durch $\tilde{F}_i = (\tilde{f}(x_{i1}, \theta), \tilde{f}(x_{i2}, \theta))$. Die Informationsmatrix kann nach Lemma 3.12 als Summe von $m = 2$ Termen dargestellt werden, und die zu verwendenden Funktionen sind die Spalten der Matrix

$$F_i^* = \tilde{F}_i V_i^{-0.5}.$$

Der Versuchsbereich \mathcal{X} sei durch $\mathcal{X} = (0, 0.04, 0.08, \dots, 2.00)$ festgelegt und enthält $r = 51$ potentielle Messstellen. Die zwei zulässigen Messungen pro Patient werden zu einem Set zusammengefasst. Doppelte Messungen an identischen Zeitpunkten sind nicht zulässig und werden daher ausgeschlossen. Der eigentliche Versuchsbereich ist dann gegeben durch $\tilde{\mathcal{X}}^2 = \{(0, 0.04), (0, 0.08), \dots, (0, 2), (0.04, 0.08), \dots, (1.96, 2.00)\}$ und enthält $r * (r - 1) / 2 = 1275$ Elemente.

Wir beginnen auch hier mit einem gleichverteilten Ausgangsdesign w^0 mit $w_i^0 = \frac{1}{1275}$ für alle $i = 1, \dots, 1275$ und führen Iterationen des Algorithmus 6.1.1 durch. Die erzielten Gewichte nach 100 Iterationen sind in Abbildung 6.3 dargestellt.

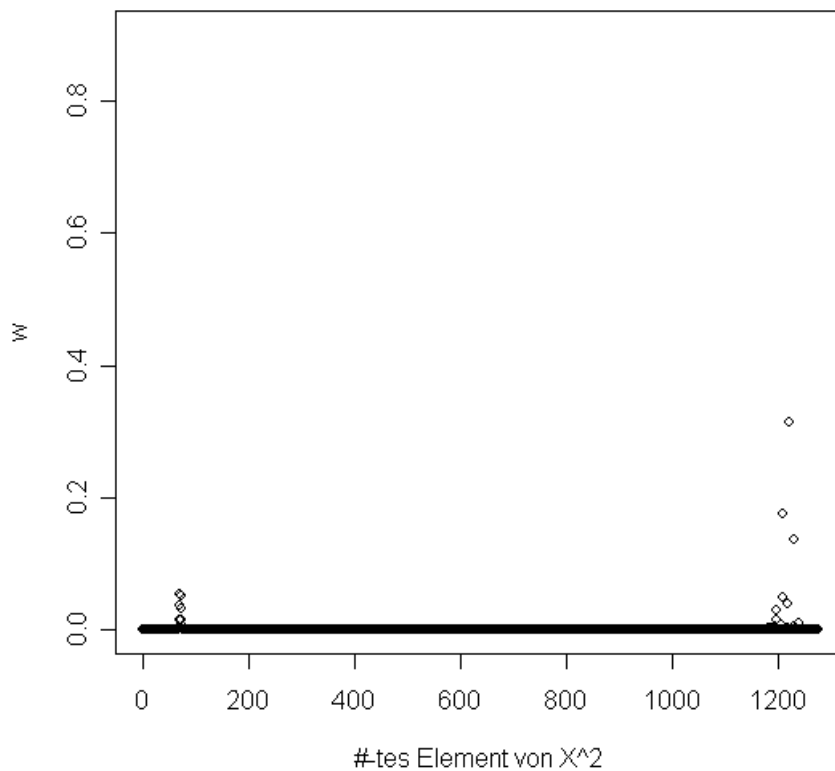


Abbildung 6.3: Verteilung der Gewichte für jedes Element von $\tilde{\mathcal{X}}^2$ nach 100 Durchläufen des Algorithmus 6.1.1 im Modell 6.1.7.

Der Algorithmus sieht vor, 20% der Beobachtungen in der Umgebung des Messsets 71 aus $\tilde{\mathcal{X}}^2$

(entspricht (0.04, 0.88)) durchzuführen und 80% in der Umgebung von Messset 1220 (entspricht (1.56, 2.00)). Auch hier kann dieses Design wieder durch das Zweipunktdesign

$$\xi = \begin{pmatrix} (0.04, 0.88) & (1.56, 2.00) \\ 0.20 & 0.80 \end{pmatrix}.$$

angenähert werden. Erneute Prüfung der Optimalität mittels Korollar 4.2 (siehe Abbildung 6.4) bestätigt auch hier die Optimalität dieses Designs. Die x -Achse von Abbildung 6.4 stellt $\tilde{\mathcal{X}}^2$ dar.

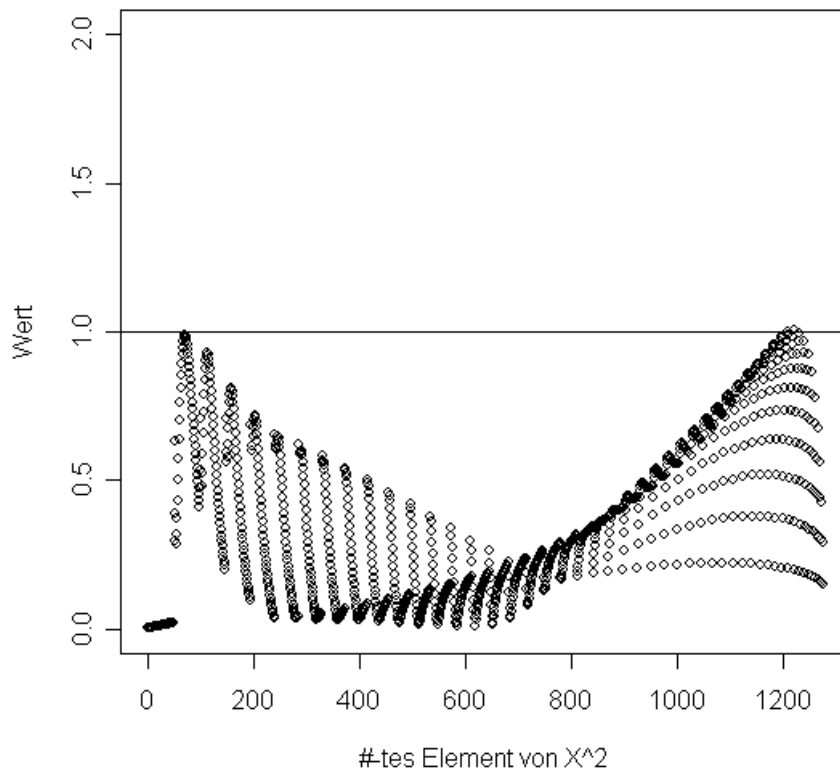


Abbildung 6.4: Darstellung des Äquivalenzsatzes 4.1.2 für das mittels Algorithmus 6.1.1 im Modell 6.1.7 erzeugte Design.

Anzumerken ist hier, dass die Größe des Versuchsbereiches bei höherem m sehr schnell zunimmt.

In der Praxis ist die Anwendung dieses Algorithmus bei Populationsmodellen daher nur bei geringem m ($m = 2$ oder $m = 3$) gut durchführbar.

6.1.3 Beispiel: ML-Schätzer bei korrelierten Beobachtungen

Als letztes Beispiel betrachten wir die Situation des ML-Schätzers im Populationsmodell für die einfache Evasion, wobei die Versuchseinheit wiederum der Satz aller m Messungen eines Probanden sei. Hier ist die Information eines m -Punkt Messsatzes an einem Probanden unter Versuchsbedingungen $x_i = (x_{i1}, \dots, x_{im})$ gegeben durch (vgl 3.2.5)

$$A_i = \tilde{F}_i V_i^{-1} \tilde{F}_i^T + \frac{1}{2} \left\{ \begin{array}{ccc} \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \dots & \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_1} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}) \\ \vdots & & \vdots \\ \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_1}) & \dots & \text{tr}(V_i^{-1} \frac{\partial V_i}{\partial \theta_k} V_i^{-1} \frac{\partial V_i}{\partial \theta_k}) \end{array} \right\}, \quad (6.1.8)$$

und die Gesamtinformation eines approximativen Designs über \mathcal{X}^m durch

$$M(\xi) = \sum_{i=1}^p A_i \xi(x_i). \quad (6.1.9)$$

Die hier zu verwendenden Gewichte sind (für $\lambda = 1$) durch die Vorschrift (6.1.5) gegeben.

Betrachten wir jetzt wieder Modell 6.1.6, diesmal aber für den ML-Schätzer im Populationsmodell.

Es gilt:

$$V_i = \tilde{F}_i^T \Omega \tilde{F}_i + \sigma^2 I_m. \quad (6.1.10)$$

Durch Einsetzen erhält man für die j, l -Elemente von V_i :

$$[V_i]_{j,l=1}^k = \frac{\partial \eta(x_i, \theta)^T}{\partial \theta_j} \Omega \frac{\partial \eta(x_i, \theta)}{\partial \theta_l} + \sigma^2 \delta_{j,l} \quad (6.1.11)$$

und für die zugehörigen Ableitungen

$$\frac{\partial [V_i]_{j,l=1}^k}{\partial \theta_r} = \frac{\partial^2 \eta(x_i, \theta)^T}{\partial \theta_j \partial \theta_r} \Omega \frac{\partial \eta(x_i, \theta)}{\partial \theta_l} + \frac{\partial \eta(x_i, \theta)^T}{\partial \theta_j} \Omega \frac{\partial \eta(x_i, \theta)}{\partial \theta_l \partial \theta_r} \quad (6.1.12)$$

Notwendig sind also die ersten und zweiten Ableitungen der Konzentrationsfunktion, die in (5.3.3) bereits bestimmt wurden.

Als Parameter seien jetzt die Populationsvarianzen $\Omega = \text{diag}(1, 0.1)$, $\sigma^2 = 1$ und zwei Beobachtungen pro Patient vorgegeben. Alle anderen Parameter bleiben wie in Beispiel 6.1.1.

Man erhält als Design das Einpunkt-Design

$$\xi = \begin{pmatrix} (0.00, 1.20) \\ 1.00 \end{pmatrix}$$

d.h. alle Patienten werden an den 2 gleichen Messpunkten gemessen. Die Abstände der Messpunkte sind dabei etwas größer als im unkorrelierten Modell.

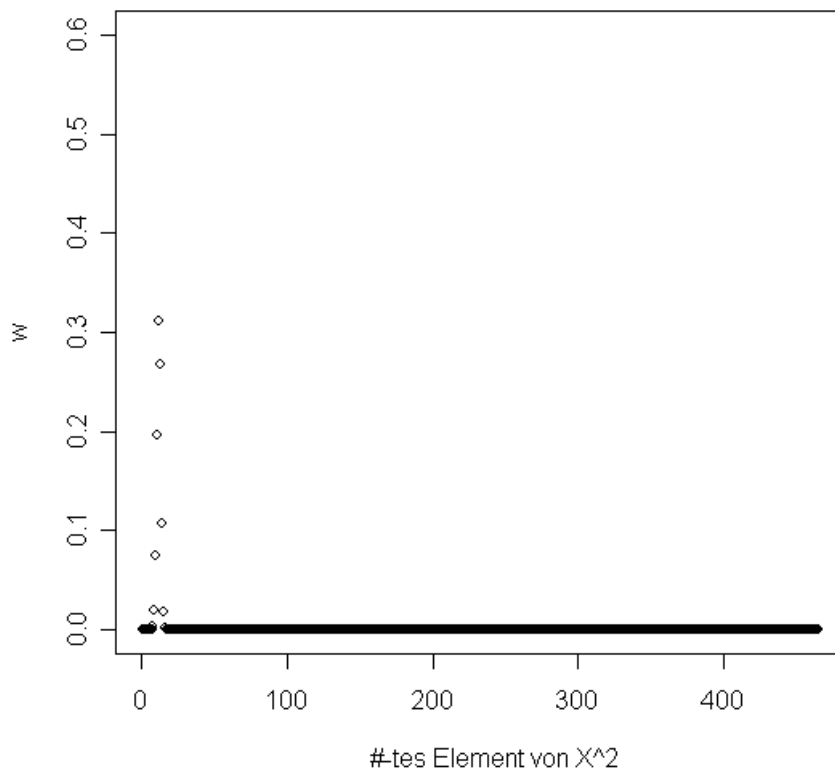


Abbildung 6.5: Verteilung der Gewichte nach 100 Durchläufen des Algorithmus 6.1.1 für den ML-Schätzer im Populationsmodell.

Kapitel 7

Asymptotisch optimale Designs bei korrelierten Beobachtungen

Wie in Kapitel 3 beschrieben existieren in der Pharmakokinetik mehrere denkbare Situationen von Modellen mit korrelierten Beobachtungen, sei es durch korrelierte Störgrößen bei Messungen an demselben Patient, oder in Populationsmodellen durch gemeinsame Realisationen der zufälligen Effekte. In solchen Situationen ist es in der Regel nicht möglich, allgemeingültige Informationsmatrizen für approximative Designs zu formulieren. Dementsprechend stehen auch die darauf basierenden Äquivalenzsätze zur Bestätigung optimaler Designs nicht zur Verfügung. Im Falle von Populationsmodellen ist es möglich, bei einer geringen Zahl von Beobachtungen pro Individuum das Problem durch Transformation des Versuchsbereiches (siehe Abschnitt 3.5.1) zu umgehen, bei vielen wiederholten Beobachtungen ist dieser Ansatz aber ebenfalls nicht mehr praktikabel. Ziel ist es jedoch, auch in solchen Situationen effiziente Designs zu finden.

Die Varianz eines gegebenen exakten Designs mit n korrelierten Beobachtungen und parameterunabhängiger Varianz ist für die beschriebenen Schätzmethoden bekannt (3.1.10 und 3.1.9). Demnach ist es bei gegebenem n möglich, Funktionale dieser Matrizen mittels bekannter Maximierungsalgorithmen zu maximieren/minimieren (z.B. Nelder-Mead oder L-BGFS Algorithmus, siehe Nelder und Mead (1965), Liu und Nocedal (1989)). Dieser Ansatz hat jedoch zwei Nachteile, zum einen muss diese Optimierung für jedes gegebene n neu durchgeführt werden,

und zum anderen ist das numerische Problem als solches sehr aufwändig. In dieser Arbeit wird daher für den NOLS-Schätzer als Alternative die Bestimmung asymptotisch optimaler Designs betrachtet, basierend auf Ideen von Bickel und Herzberg (1979). Auch diese erfordern die Anwendung von Optimierungsalgorithmen, allerdings für numerisch einfachere Probleme. Wir beginnen mit der Beschreibung eines Vertreters klassischer Maximierungsalgorithmen. Die eigentlichen asymptotisch optimalen Designs werden im zweiten Abschnitt betrachtet.

7.1 Der Nelder-Mead Algorithmus

Der Nelder-Mead Algorithmus (Nelder und Mead (1965)) erlaubt die Minimierung einer gegebenen reellwertigen Funktion g bezüglich eines s -dimensionalen Parametervektors $x \in \mathcal{X}^s \subset \mathbb{R}^s$. Zu diesem Zweck müssen zunächst $s + 1$ verschiedene Elemente von \mathcal{X}^s als Startwerte vorgegeben werden, geordnet bezüglich der Größe ihrer Funktionswerte. Der $(s + 1)$ -te Punkt ist der ungünstigste und wird in jedem Durchgang des Algorithmus durch potentiell bessere Punkte ersetzt. Der eigentliche Ablauf jedes Durchgangs ist der folgende:

1. Ordne die Punkte bezüglich der Größe ihrer Funktionswerte. Vergebe Bezeichnungen b_1, \dots, b_{s+1} , so dass gilt

$$g(b_1) \leq g(b_2) \leq \dots \leq g(b_{s+1}). \quad (7.1.1)$$

2. Bestimme den Mittelwert b_0 der s besten Punkte b_1 bis b_s .
3. Spiegelung: Erhalte einen neuen Punkt r durch Spiegelung von b_{s+1} am Punkt b_0 mittels der Vorschrift $r = b_0 + \alpha * (b_0 - b_{s+1})$. Die Steuerkonstante α wird üblicherweise als $\alpha = 1$ vorgegeben.

Falls $g(b_1) \leq g(r) < g(b_{s+1})$ ersetze b_{s+1} durch r und gehe zu Schritt 1.

Falls $g(r) < g(b_1)$ gehe zu Schritt 4.

Ansonsten gehe zu Schritt 5.

4. Expansion: Erhalte einen neuen Punkt e mittels der Vorschrift $e = b_0 + \gamma * (b_0 - b_{s+1})$. Die Steuerkonstante γ wird üblicherweise als $\gamma = 2$ vorgegeben.
Falls $g(e) < g(r)$ ersetze b_{s+1} durch e und gehe zu Schritt 1.
Ansonsten ersetze b_{s+1} durch r und gehe ebenfalls zu Schritt 1.
5. Kontraktion: Erhalte einen neuen Punkt c mittels der Vorschrift $c = b_0 + \rho * (b_0 - b_{s+1})$. Die Steuerkonstante ρ wird üblicherweise als $\rho = 0.5$ vorgegeben.
Falls $g(c) < g(b_{s+1})$ ersetze b_{s+1} durch c und gehe zu Schritt 1.
Ansonsten gehe zu Schritt 6.
6. Reduktion: Ersetze alle außer dem besten Punkt durch die Vorschrift $b_i^{neu} = b_1 + \sigma * (b_i - b_1)$, $i = 2, \dots, (s + 1)$. Die Steuerkonstante σ wird üblicherweise als $\sigma = 0.5$ vorgegeben.
Gehe zu Schritt 1.

Der Nelder-Mead Algorithmus konvergiert langsam, ist aber robust gegenüber den Modellvorgaben (Walters et al. (1991)).

7.2 Asymptotisch effiziente Designs

Als Modell betrachten wir ein linearisiertes Populationsmodell mit zufälligen Effekten

$$Y_{ij} \approx \eta(x_{ij}, \theta) + \tilde{f}(x_{ij}, \theta)^T (b_i - \theta) + \epsilon_{ij} \quad i = 1, \dots, N, j = 1, \dots, m. \quad (7.2.1)$$

Es bezeichne Y_{ij} die Beobachtungen der j -ten Messung am i -ten Patienten, $x_{ij} \in \mathcal{X}$ die zugehörige Versuchsbedingung, $b_i = (b_{i1}, \dots, b_{ik})^T \in \Theta \subset \mathbb{R}^k$ die Realisationen der Parameter des i -ten Patienten. Weiterhin sei $\eta : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ eine bekannte, zweimal differenzierbare und nicht notwendig lineare Funktion, $\tilde{f}(x_{ij}, \theta)$ der Spaltenvektor der in Abschnitt 3.2 definierten Gradientenfunktion und $\tilde{F}_i = (\tilde{f}(x_{i1}, \theta), \dots, \tilde{f}(x_{im}, \theta))$.

Die Störgrößen ϵ_{ij} seien normalverteilt mit Erwartungswert 0 und Varianz σ^2 . Beobachtungen an unterschiedlichen Patienten seien unkorreliert und für die Kovarianz bei Beobachtungen am selben Patienten gelte:

$$\text{Cov}(\epsilon_{ij}, \epsilon_{il}) = \sigma^2(\gamma r(x_{ij} - x_{il}) + (1 - \gamma)\delta_{j,l}), \quad (7.2.2)$$

wobei $\gamma \in [0, 1]$, $r(u)$ sei eine vorgegebene Korrelationsfunktion mit $r(0) = 1$ und $\delta_{j,l}$ bezeichne Kronecker's Symbol. Die gesamte Kovarianzmatrix der Störgrößen bei Beobachtungen am Patienten i sei als W_i bezeichnet. Die zufälligen Effekte b_i sind Realisierungen von unabhängigen Zufallsvariablen einer beliebigen Verteilung mit Erwartungswert $\theta = (\theta_1, \dots, \theta_k)$ und Varianz Ω . Ausserdem seien diese Zufallsvariablen unabhängig von den Störgrößen ϵ_{ij} . Die Kovarianz zweier Beobachtungen unter Versuchsbedingungen x_{ij} und x_{il} ($j \neq l$) ist daher

$$\text{Cov}(Y_{ij}, Y_{il}) = \tilde{f}^T(x_{ij}, \theta)\Omega\tilde{f}(x_{il}, \theta) + \sigma^2\gamma r(x_{ij} - x_{il}),$$

und die gesamte Kovarianzmatrix von $(Y_{i1}, \dots, Y_{im})^T$ ist

$$V_i := \tilde{F}_i^T\Omega\tilde{F}_i + W_i, \quad i = 1, \dots, N. \quad (7.2.3)$$

Wir nehmen an, dass die Korrelationsfunktion $r(u)$ von m abhängt in der Form

$$r(u) = r_m(u) = \rho(mu) \quad (7.2.4)$$

für eine gleichbleibende Funktion ρ mit $\rho(u) \rightarrow 0$ für $u \rightarrow \infty$. Dies erlaubt die Konstruktion einer einheitlichen Designdichte für alle m . Ohne diese Voraussetzung muss für jedes m eine eigene Designdichte konstruiert werden, das Grundprinzip bleibt aber unverändert.

Nach Schmelter (2007) verwendet ein optimales Design für Populationsmodelle asymptotisch die selbe Kombination von m Messungen für alle Patienten, d.h. ohne Einschränkung gelte $x_{ij} = x_j$ ($i = 1, \dots, N, j = 1, \dots, m$). Entsprechend reduziert sich das hier beschriebene Problem auf die Bestimmung eines einzelnen Designvektors $(x_1, \dots, x_m)^T$. Ebenso sind dann die Matrizen \tilde{F}_i , W_i und V_i identisch für alle $i = 1, \dots, N$ und werden als \tilde{F} , W und V bezeichnet. Wir gehen ohne Beschränkung der Allgemeinheit davon aus, dass der Vektor der Versuchsbedingungen der Größe nach geordnet ist, d.h. $x_1 < \dots < x_m$.

Wir verwenden in dieser Situation zur Schätzung den NOLS-Schätzer. Die Varianzmatrix einer Beobachtungsreihe unter Bedingungen $x = (x_1, \dots, x_m)^T$ ist dann wie in (3.1.9)

$$\text{Var}(\hat{\theta}_{NOLS}) \approx (\tilde{F}\tilde{F}^T)^{-1}\tilde{F}V\tilde{F}^T(\tilde{F}\tilde{F}^T)^{-1}. \quad (7.2.5)$$

Grundprinzip der asymptotisch optimalen Designs ist es, exakte n -Punkt Designs in dieser Situation durch eine stetige Designdichte über den gesamten Versuchsbereich zu approximieren. Identifiziert wird diese durch eine Quantilfunktion $a : [0, 1] \rightarrow \mathcal{X}$. Ein konkretes n -Punkt Design kann daraus dann durch

$$x_j = a\left(\frac{j-1}{n-1}\right) \quad j = 1, \dots, m \quad (7.2.6)$$

bestimmt werden. Nimmt man den Versuchsbereich als $\mathcal{X} = [0, 2]$ an, so beschreibt als Beispiel $a(u) = 2u$ ein gleichverteiltes Design auf $[0, 2]$. Im Fall $m = 5$ wären die Messpunkte gerade $x = (0, 0.5, 1, 1.5, 2)$.

Ziel ist es jetzt, die Varianzmatrix einer solchen asymptotischen Designdichte herzuleiten und über den Raum der Designdichten zu minimieren. Zu diesem Zweck bezeichnen wir die Komponenten des Vektors $\tilde{f}(u, \theta)$ mit $\tilde{f}_1(u, \theta), \dots, \tilde{f}_k(u, \theta)$ und treffen die folgenden Regularitätsannahmen:

(C1) Die Regressionsfunktionen $\tilde{f}_1(u, \theta), \dots, \tilde{f}_k(u, \theta)$ sind linear unabhängig, beschränkt auf dem Versuchsbereich \mathcal{X} und Lipschitz stetig, d.h.

$$|f_i(t, \theta) - f_i(s, \theta)| \leq K|t - s| \quad \text{und} \quad |f_i(t, \theta)| \leq K \quad \text{for all } t, s \in \mathcal{X}, i = 1, \dots, k, K > 0.$$

(C2) Die Funktion a ist zweifach differenzierbar und es existiert eine positive Konstante $K < \infty$ so dass für alle $u \in (0, 1)$

$$\frac{1}{K} \leq \frac{da(u)}{du} \leq K, \quad \left| \frac{d^2a(u)}{du^2} \right| \leq K. \quad (7.2.7)$$

(C3) Die Korrelationsfunktion ρ ist differenzierbar mit beschränkter Ableitung und diese Ableitung ist ≤ 0 für ausreichend große Argumente.

Annahme (C1) beschreibt die Stetigkeit der Linkfunktionen als Funktionen der Versuchsbedingungen und ist in den meisten gängigen Regressionsfunktionen und speziell den hier verwendeten Exponentialfunktionen gegeben. Auch die meisten übliche Korrelationsfunktionen, speziell auch die exponentiell abfallende Korrelation erfüllen auf kompakten Intervallen Bedingung (C3). Bedingung (C2) schließt zeitlich extrem konzentrierte Messungen aus, die allein aus praktischen

Erwägungen bei wiederholten Messungen um selben Patienten nicht durchführbar sind.

Unter diesen Annahmen kann die Informationsmatrix einer asymptotischen Designdichte mittels des folgenden Lemmas dargestellt werden:

Lemma 7.1

Seien Bedingungen (C1)-(C3) erfüllt. Dann gilt:

$$m^{-1} \sum_{i=1}^m \sum_{j \neq i}^m \tilde{f}_k(x_i, \theta) \tilde{f}_l(x_j, \theta) \rho(m(x_j - x_i)) = 2 \int_0^1 \tilde{f}_k(a(u), \theta) \tilde{f}_l(a(u), \theta) Q\left(\frac{da(u)}{du}\right) du + o(1),$$

falls dieser Term endlich ist. Die Funktion Q ist definiert als $Q(u) = \sum_{i=1}^{\infty} \rho(iu)$.

Beweis

Siehe Bickel und Herzberg (1979). \square

Lemma 7.2

Bedingungen (C1), (C2) and (C3) seien erfüllt. Dann gilt für die Kovarianzmatrix des nichtlinearen gewöhnliche kleinste Quadrate Schätzers:

$$Var(\hat{\theta}_{NOLS}) = \frac{\sigma^2}{m} (P^{-1}(a) + 2\gamma P^{-1}(a)R(a)P^{-1}(a)) + \Omega + o\left(\frac{1}{m}\right), \quad (7.2.8)$$

wobei die Matrizen P und R definiert sind durch

$$P(a) = \left(\int_0^1 \tilde{f}_j(a(u), \theta) \tilde{f}_l(a(u), \theta) du \right)_{j,l=1}^k,$$

$$R(a) = \left(\int_0^1 \tilde{f}_j(a(u), \theta) \tilde{f}_l(a(u), \theta) Q\left(\frac{da(u)}{du}\right) du \right)_{j,l=1}^k,$$

und die Funktion Q durch

$$Q(u) = \sum_{i=1}^{\infty} \rho(iu). \quad (7.2.9)$$

Beweis

Es gilt:

$$Var(\hat{\theta}_{NOLS}) \approx (\tilde{F}\tilde{F}^T)^{-1} \tilde{F}V\tilde{F}^T (\tilde{F}\tilde{F}^T)^{-1} \quad (7.2.10)$$

$$\begin{aligned}
&= (\tilde{F}\tilde{F}^T)^{-1}\tilde{F}(W + \tilde{F}^T\Omega\tilde{F})\tilde{F}^T(\tilde{F}\tilde{F}^T)^{-1} \\
&= (\tilde{F}\tilde{F}^T)^{-1}\tilde{F}W\tilde{F}^T(\tilde{F}\tilde{F}^T)^{-1} + \Omega
\end{aligned}$$

Für die Komponenten ist bekannt, dass

$$(\tilde{F}\tilde{F}^T)_{j,l} = \sum_{r=1}^m \tilde{f}_j(x_r, \theta)\tilde{f}_l(x_r, \theta) \quad j, l = 1, \dots, k$$

und

$$(\tilde{F}W\tilde{F}^T)_{j,l} = \sigma^2(\gamma \sum_{r=1}^m \sum_{s \neq r}^m \tilde{f}_j(x_r, \theta)\tilde{f}_l(x_s, \theta)\rho(m(x_s - x_r)) + \sum_{r=1}^m \tilde{f}_j(x_r, \theta)\tilde{f}_l(x_r, \theta)) \quad j, l = 1, \dots, k.$$

Die Behauptung folgt dann aus Lemma 7.1. \square

Substituiert man nun die Quantilfunktion a durch die zugehörige Dichtefunktion $\varphi : \mathcal{X} \rightarrow [0, 1]$ so hat man $\frac{da(u)}{du} = \frac{1}{\varphi(u)}$ und die Kovarianz des OLS-Schätzers kann für große m angenähert werden durch $U(\varphi)/m$, wobei die $k \times k$ Matrix U gegeben ist durch

$$U(\varphi) := \sigma^2 (P^{-1}(\varphi) + 2\gamma P^{-1}(\varphi)R(\varphi)P^{-1}(\varphi)) + \Omega. \quad (7.2.11)$$

Die Matrizen P und R sind jetzt gegeben durch

$$\begin{aligned}
P(\varphi) &= \left(\int_{\mathcal{X}} \tilde{f}_i(u, \theta)\tilde{f}_j(u, \theta)\varphi(u) du \right)_{i,j=1}^k, \\
R(\varphi) &= \left(\int_{\mathcal{X}} \tilde{f}_i(u, \theta)\tilde{f}_j(u, \theta)Q(1/\varphi(u))\varphi(u) du \right)_{i,j=1}^k.
\end{aligned}$$

Optimale asymptotische Designdichten minimieren geeignete Funktionale der Matrix $U(\varphi)$. In diesem Zusammenhang liegt der Fokus wiederum auf c -optimalen Designs, d.h. Designs die $c^T U(\varphi) c$ minimieren ($c \in \mathbb{R}^k$). Es gilt

$$c^T U(\varphi) c = c^T \sigma^2 (P^{-1}(\varphi) + 2\gamma P^{-1}(\varphi)R(\varphi)P^{-1}(\varphi)) c + c^T \Omega c,$$

d.h. es reicht aus,

$$c^T (P^{-1}(\varphi) + 2\gamma P^{-1}(\varphi)R(\varphi)P^{-1}(\varphi)) c \quad (7.2.12)$$

zu minimieren.

Nächster Schritt ist es jetzt, eine optimale Designdichte numerisch zu bestimmen. In dieser Arbeit wird die Dichte in der folgenden Weise als eine polynomielle Repräsentation dargestellt:

$$\varphi(u) = \frac{(p_0 + p_1u + \dots + p_ru^r)_+}{\int_{\mathcal{X}}(p_0 + p_1u + \dots + p_ru^r)_+ du}$$

Die mit + markierten Terme werden falls negativ auf null gesetzt. Diese Dichte kann damit notwendigerweise ausschließlich Werte ≥ 0 annehmen. Notwendig ist jetzt, Parameter p_1, \dots, p_r zu ermitteln, die 7.2.12 minimieren. Dies kann mit Hilfe gängiger Maximierungsalgorithmen geschehen. In dieser Arbeit wird dazu ein Polynom 6. Ordnung ($r = 6$) und der Nelder-Mead Algorithmus verwendet (siehe 7.1).

Designs basierend auf asymptotisch optimalen Designdichten sind nicht zwingend auch im Fall endlicher Versuchsumfänge optimal. Es ist daher sinnvoll, die Qualität dieser Designs wenn möglich mit tatsächlich optimalen Designs zu vergleichen. Man erhält das Konzept der Effizienz eines Designs.

Definition 7.3 Sei $c \in \mathbb{R}^k$ und $\xi_{Opt} \in \Xi$ ein c -optimales Design zur Schätzung der Parameterkombination $c^T \theta$ und $A(\xi)$ die Varianzmatrix einer Schätzung von θ durch das Design $\xi \in \Xi$. Die Effizienz eines Designs $\xi \in \Xi$ ist gegeben durch

$$Eff(\xi) = \frac{c^T A(\xi_{Opt})c}{c^T A(\xi)c}. \quad (7.2.13)$$

7.2.1 Beispiele

Der medizinische Hintergrund der Beispiele dieses Abschnittes ist eine reale Pilotstudie an 4 Patienten (Thürmann et al. (2004)). In dieser war beobachtet worden, dass Patienten, die das frei verkäufliche pflanzliche Durchfall-Präparat Uzara® einnahmen (als Tropfen, d.h. in oraler Verabreichung) in klinischen Bluttests stark erhöhte Werte für die Serumkonzentrationen des

Herzinsuffizienzmedikamentes Digitoxin aufwiesen. Ursache dafür ist die chemische Ähnlichkeit der beiden Substanzen. Da Uzara jedoch nicht herzwirksam ist, kann dies zu erheblichen Problemen bei der Steuerung von Behandlungsprogrammen zur Herzinsuffizienz führen. Es wurde daher entschieden, die pharmakokinetischen Eigenschaften einer oralen Gabe von Uzara mit den Eigenschaften der üblichen intravenösen Gabe einer regulären Digitoxin Medikation (Lanicor®) anhand einer größeren Stichprobe von 18 Patienten zu vergleichen, jeweils mit wiederholten Messungen der Serumkonzentrationen im Verlauf der nächsten 36 Stunden. Das Hauptaugenmerk lag dabei auf dem Vergleich der Fläche unter der Kurven als Maß für den Gesamteffekt der Behandlung, das recht gut auch zwischen verschiedenen Verabreichungsformen zu vergleichen ist.

Wir beginnen mit der Situation bei der Gabe von Lanicor, d.h. wir betrachten wieder das Einkompartimentmodell mit intravenöser Verabreichung, in der Form 7.2.1. Die zugrundeliegende Konzentrationsfunktion ist $\eta(x, \theta) = \theta_1 e^{-\theta_2 x}$ mit $\theta = (\theta_1, \theta_2)^T$, d.h. es gilt $\tilde{f}(x, \theta) = (e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x})^T$. Die Korrelationsfunktion sei durch $\rho(u) = e^{-\lambda u}$ gegeben. Für die Parameter gelte (nach Schätzung aus der Pilotstudie) $\hat{\theta} = (30, 0.75)^T$,

$$\hat{\Omega} = \begin{pmatrix} 9 & 0.189 \\ 0.189 & 0.0049 \end{pmatrix}$$

sowie $\hat{\gamma} = 0.8$, $\hat{\lambda} = 0.05$ und $\hat{\sigma}^2 = 0.6$. Der Designbereich ist auf $\mathcal{X} = [0, 36]$ festgelegt. Insgesamt stehen 14 Beobachtungen zur Verfügung, und es soll die Fläche unter der Kurve, d.h. $c^T \theta$ für den c -Vektor $c = \frac{1}{\theta_2} (1, -\frac{\theta_1}{\theta_2})^T$ geschätzt werden.

Die resultierende Designdichte ist in Abbildung 7.1 dargestellt.

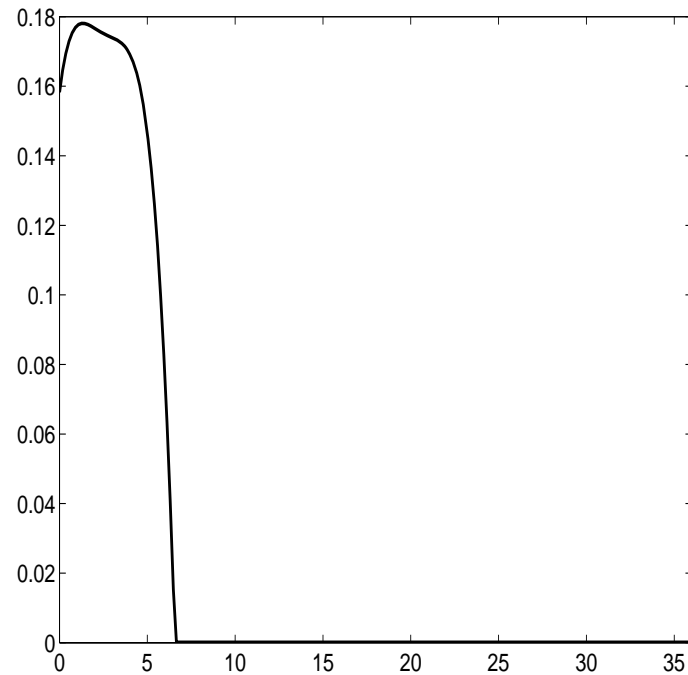


Abbildung 7.1: *Asymptotische Designdichte zur Schätzung Der Fläche unter der Kurve im Modell des intravenösen Einkompartiment-Modells.*

Anhand der entsprechenden Quantile dieser Verteilung ($a(\frac{0}{14-1}), a(\frac{1}{14-1}), \dots, a(\frac{13}{14-1})$) erhält man das exakte 14-Punkt-Design

$$(0.432, 0.85, 1.25, 1.66, 2.06, 2.46, 2.87, 3.28, 3.7, 4.12, 4.55, 5.01, 5.54, 36).$$

Direkte numerische Maximierung von 7.2.5 über den Versuchsbereich der 14-Punkt Designs liefert das Design

$$(0.75, 0.84, 0.94, 1.06, 1.21, 1.38, 1.58, 1.83, 2.12, 2.47, 2.89, 3.41, 4.07, 36.0).$$

Die Effizienz des asymptotischen Designs beträgt 0.95. Vergleicht man im Gegensatz dazu ein naives, gleichverteiltes Design

$$(0.0, 2.8, 5.5, 8.3, 11.1, 13.9, 16.6, 19.4, 22.2, 24.9, 27.7, 30.5, 33.2, 36.0)$$

so beträgt die Effizienz lediglich 0.42. Asymptotisch optimale Designdichten erlauben es also auch bei endlichen Stichprobengrößen wenn auch nicht optimale, so doch effiziente Designs zu

bestimmen.

Als zweites Beispiel betrachten wir nun die Situation bei oraler Gabe von Uzara, das heißt im Einkompartiment Modell mit oraler Verabreichung. Hier ist die Konzentrationsfunktion also die Bateman-Funktion $\eta(x, \theta) = \theta_1(e^{-\theta_2 x} - e^{-\theta_3 x})$ mit den Parametern $\theta = (\theta_1, \theta_2, \theta_3)^T$, und $\tilde{f}(x, \theta) = (e^{-\theta_2 x} - e^{-\theta_3 x}, -\theta_1 x e^{-\theta_2 x}, \theta_1 x e^{-\theta_3 x})^T$. Die Korrelationsfunktion bleibe $\rho(u) = e^{-\lambda u}$ und der Versuchsbereich $\mathcal{X} = [0, 36]$. Die Parameter seien $\theta = (0.2, 0.135, 28)^T$,

$$V_p = \begin{pmatrix} 0.0025 & 0.0019 & 0 \\ 0.0019 & 0.0016 & 0 \\ 0 & 0 & 144 \end{pmatrix}$$

und $\gamma = 0.8$, $\lambda = 0.01$, $\sigma^2 = 0.2$. Auch hier ist die AUC zu schätzen, d.h. $c = (\frac{1}{\theta_2} - \frac{1}{\theta_3}, -\frac{\theta_1}{\theta_2^2}, \frac{\theta_1}{\theta_3^2})$.

Insgesamt stehen diesmal 15 Beobachtungen zur Verfügung.

Die resultierende Designdichte ist in Abbildung 7.2 dargestellt.

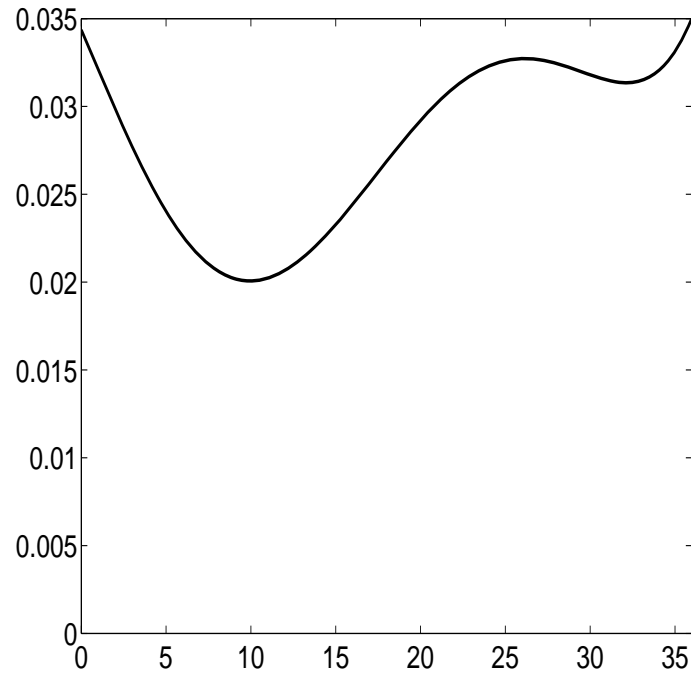


Abbildung 7.2: *Asymptotische Designdichte zur Schätzung der Fläche unter der Kurve im Bateman-Modell.*

Man erhält man das exakte 15-Punkt-Design

(2.09, 4.55, 7.49, 10.8, 13.9, 16.8, 19.2, 21.5, 23.6, 25.7, 27.7, 29.8, 31.9, 34, 36).

Von diesem Design ausgehende numerische Maximierung liefert das Design

(2.47, 3.07, 4.11, 12.37, 14.52, 16.38, 18.20, 20.07, 22.03, 24.11, 26.36, 28.79, 31.46, 34.39, 36.0).

Die Effizienz des asymptotischen Designs beträgt 0.99. Beide Designs sind sehr ähnlich zum gleichverteilten Design

(0.00, 2.57, 5.14, 7.71, 10.28, 12.85, 15.42, 17.99, 20.56, 23.13, 25.70, 28.27, 30.84, 33.41, 35.98),

dessen Effizienz 0.97 beträgt. Das naive Design ist in dieser Situation also bereits nahezu optimal.

Kapitel 8

Zusammenfassung

Ziel dieser Arbeit war, Methoden optimaler Versuchsplanung auf Problemstellungen klinischer Pharmakokinetik zu übertragen. Die dort verwendeten Modelle mit zufälligen Effekten resultieren in zwei erschwerenden Eigenschaften. Zum einen hängt zusätzlich zum Erwartungswert auch die Varianz der Beobachtungen von den zu schätzenden Parametern ab und zum anderen sind verschiedene Messungen am selben Patienten korreliert. In beiden Fällen ist die klassische Darstellung von Informationsmatrizen bestimmter Beobachtungen nicht mehr gegeben, und gängige Techniken der Versuchsplanung sind in der üblichen Fassung nicht mehr anwendbar. Notwendig war es also, entsprechende Verallgemeinerungen der klassischen Methoden zu entwickeln.

Für das Problem parameterabhängiger Varianz zeigte sich, dass dort die Informationsmatrix (für Maximum Likelihood Schätzungen) für jede einzelne Beobachtung als Summe von zwei Submatrizen in der üblichen Struktur dargestellt werden kann. Der bekannte Satz zu geometrischen Representationen von Versuchsplanungsproblemen (Elfving (1952)) konnte so verallgemeinert werden, das auch solche Summen von Informationsmatrizen dargestellt werden können. Auch die klassischen Äquivalenzsätze (Pukelsheim (1993)) zur Überprüfung der Optimalität eines gegebenen Design sowie gängige numerische Verfahren zur Designbestimmung konnten problemlos auf diese Situation übertragen werden.

Etwas schwieriger gestaltete sich die Situation für korrelierte Daten. Hier ist die Additivität der Informationsmatrizen verschiedener Beobachtungen nicht mehr gegeben. Demnach ist eine allge-

meine fallzahlunabhängige Darstellung der Informationsmatrix für gegebene Designs prinzipiell nicht möglich, und auch die darauf basierenden Äquivalenzsätze und geometrischen Darstellung sind nicht mehr verfügbar. In dieser reinen Situation kann direkte Designbestimmung nur noch für eine festgelegte Fallzahl durch klassische numerische Maximierungsprozeduren ermöglicht werden. Lediglich für bestimmte Spezialfälle ist es möglich, die Effizienz solcher Prozeduren durch Bestimmung asymptotisch optimaler Designdichten zu erhöhen.

Für die spezielle Korrelationsstruktur von Populationsmodellen ist es jedoch möglich, die Problematik durch Verschiebung des Designbereiches zu umgehen. Zu diesem Zweck werden alle m Messungen eines Patienten zu einem Messsatz zusammengefasst. Messsätze verschiedener Patienten sind dann wieder voneinander unabhängig und die Optimierung erfolgt über den Raum dieser m -Punkt-Messsätze. Unter bestimmten Regularitätsbedingungen konnte gezeigt werden, dass sich auch die Informationsmatrizen solcher m -Punkt Messsätze ebenfalls als Summe von Einzelmatrizen darstellen lässt, und es ist erneut möglich, graphische Darstellungen, Äquivalenzsätze und Algorithmen zu formulieren.

Symbolverzeichnis

k	Anzahl der zu schätzenden Parameter
n	Gesamtzahl der verfügbaren Beobachtungen
$\theta = (\theta_1, \dots, \theta_k)^T$	Vektor der zu schätzenden Parameter, $\theta \in \Theta \subset \mathbb{R}^k$
$x = (x_1, \dots, x_n)^T$	Vektor der Versuchsbedingungen der n verfügbaren Beobachtungen
\mathcal{X}	Menge der möglichen Versuchsbedingungen (kompakte Untermenge der reellen Zahlen)
$\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$	Vektor der zufälligen Störgrößen
$\eta(x, \theta)$	Regressionsfunktion
$Y = (Y_1, \dots, Y_n)^T$	Zufallsvektor der n Beobachtungen
m	In Populationsmodellen: Anzahl der Beobachtungen pro Patient
N	In Populationsmodellen: Anzahl der verfügbaren Patienten
b_i	In Populationsmodellen: Realisation des zufälligen Effekts für den i -ten Patienten ($i = 1, \dots, N$)
Ω	In Populationsmodellen: Varianzmatrix der zufälligen Effekte
ξ	Wahrscheinlichkeitsverteilung (Design) über den Raum der Versuchsbedingungen (\mathcal{X} oder \mathcal{X}^m)
p	Anzahl der Trägerpunkte eines Designs ξ
$M(\xi)$	Informationsmatrix des Designs ξ
\mathcal{M}	Raum der Informationsmatrizen
$M(\xi)^-$	Verallgemeinerte Inverse der Informationsmatrix (g-Inverse)

V	Varianzmatrix aller n Beobachtungen eines Versuchs
V_i	Varianzmatrix der m Beobachtungen an Individuum i , $i = 1, \dots, N$
F	Designmatrix der gesamten Versuchsreihe
F_i	Designmatrix der Messungen an Individuum i , $i = 1, \dots, N$
\tilde{F}	linearisierte Designmatrix der gesamten Versuchsreihe
\tilde{F}_i	linearisierte Designmatrix der Messungen an Individuum i , $i = 1, \dots, N$
\tilde{f}	Komponenten der linearisierten Designmatrix
\mathcal{R}	Elfving-Raum
AUC	Fläche unter der Kurve (der Konzentrationsfunktion)
$\mathcal{A}(c)$	Menge der Designs ξ , mit denen $c^T \theta$ schätzbar ist
w_i	Anteil der Messungen am Punkt x_i im Kontext multiplikativer Algorithmen, $i = 1, \dots, r$
r	Anzahl der möglichen Messpunkte im Kontext multiplikativer Algorithmen
S	Menge aller Designs mit positiven Gewichten für alle Messpunkte im Kontext multiplikativer Algorithmen
$\text{diag}(A)$ bzw.	Diagonalmatrix deren Diagonaleinträge den Diagonalelementen
$\text{diag}(a)$	der Matrix A bzw. den Einträgen des Vektors a entsprechen
$o(1)$	Landau Symbol, Term geht gegen null für $m \rightarrow \infty$
$\langle A, B \rangle$	Skalarprodukt der Matrizen A und B , definiert als $\text{tr}(AB)$
$\text{vec}(A)$	Die Spalten einer $n \times k$ -Matrix, geschrieben als Spaltenvektor der Länge nk
$\delta_{j,l}$	Kronecker Symbol, 1 falls $j = l$, 0 sonst

Literaturverzeichnis

- Atkinson, A. C., Chaloner, K., Herzberg, A. M., und Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics*, 49:325–337.
- Beatty, D. A. und Piegorsch, W. W. (1997). Optimal statistical design for toxicokinetic studies. *Statistical Methods in Medical Research*, 6:359–376.
- Bickel, P. J. und Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Ann. Statist.*, 7(1):77–95.
- Cayen, M. und Black, H. (1993). *Role of toxicokinetics in dose selection for carcinogenicity studies*. Welling, P., de la Iglesia, F: Drug toxicokinetics. Marcel Dekker, New York, N. Y.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.*, 24:586–602.
- Cox, D. und Hinkley, D. (2000). *Theoretical Statistics*. Chapman and Hall, Boca Raton.
- Dette, H., Bretz, F., Pepelyshev, A., und Pinheiro, J. C. (2008a). Optimal designs for dose finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.
- Dette, H. und Holland-Letz, T. (2009). A geometric characterization of c -optimal designs for heteroscedastic regression. *Ann. Statist.*, 37(6B):4088–4103.
- Dette, H., Pepelyshev, A., und Zhigljavsky (2008b). Improving updating rules in multiplicative algorithms for computing d -optimal designs. *Computational Statistics and Data Analysis*, 53:312–320.
- Elfving, G. (1952). Optimal allocation in linear regression theory. *Annals of Mathematical Statistics*, 23:255–262.

- Ette, E., Kelman, A., Howie, C., und Whiting, B. (1995). Analysis of animal pharmacokinetic data: Performance of the one point per animal design. *Journal of Pharmacokinetics and Biopharmaceutics*, 23:551–566.
- Holland-Letz, T., Dette, H., und Pepelyshev, A. (2009). A geometric characterization of c -optimal designs for regression models with correlated observations. *Submitted for publication*.
- Jones, B., Wang, J., Jarvis, P., und Byrom, W. (1999). Design of crossover trials for pharmacokinetic studies. *Journal of Statistical Planning and Inference*, 78:307–316.
- Karlin, S. und Studden, W. (1966). *Tchebycheff Systems with Applications in Analysis and Statistics*. Wiley, New York.
- Kiefer, J. (1974). General equivalence theory for optimum designs. *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.
- Källén, A. (2008). *Computational Pharmacokinetics*. Chapman and Hall, Boca Raton.
- Lehmann, E. und Casella, G. (1998). *Theory of Point Estimation*. Springer, New York.
- Liu, D. und Nocedal, J. (1989). On the limited memory method for large scale optimization. *JMathematical Programming B.*, 45.
- Magnus, J. R. und Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, Chichester.
- Nelder, J. und Mead, R. (1965). A simplex method for function minimization. *Computer Journal.*, 7:308–313.
- Pázman, A. (1986). *Foundations of Optimum Experimental Design*. D. Reidel Publishing Company, Dordrecht.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. John Wiley & Sons, New York.
- Pukelsheim, F. und Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, 79:763–770.

- Retout, S., Duffull, S., und Mentre, F. (2001). Development and implementation of the Fisher information matrix for the evaluation of population pharmacokinetic designs. *Computer Methods and Programs in Biomedicine*, 65:141–151.
- Retout, S. und Mentré, F. (2003). Further developments of the Fisher information matrix in nonlinear mixed-effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 13:209–227.
- Rowland, M. (1993). *Clinical Pharmacokinetics: Concepts and Applications*. Williams and Wilkins, Baltimore.
- Schmelter, T. (2007). The optimality of single-group designs for certain mixed models. *Metrika*, 65(2):183–193.
- Searle, S. (1971). *Linear Models*. John Wiley, New York.
- Searle, S. R. (1982). *Matrix Algebra useful for Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley and Sons, New York.
- Shargel, L. (1993). *Applied Biopharmaceutics and Pharmacokinetics*. Chapman and Hall, London.
- Sheiner, W. und Wakefield, J. (1999). Population modelling in drug development. *Statistical Methods in Medical Research*, 8:183–193.
- Thürmann, P., Neff, A., und Fleisch, J. (2004). Interference of Uzara glycosides in assays of digitalis glycosides. *International Journal of Clinical Pharmacology and Therapeutics*, 42(5):281–284.
- Walters, F., Parker, L., Morgan, S., und Deming, S. (1991). *Sequential Simplex Optimization*. CRC Press, Boca Raton.
- Yu, Y. (2009). Monotonic convergence of a general algorithm for computing optimal designs. *To appear in: Annals of Statistics*.