

# Maximum-Likelihood- Approximation mit dünnen Gittern

Fabian Schildmann

Geboren am 24. Februar 1990 in Bielefeld

8. August 2012

Bachelorarbeit Mathematik

Betreuer: Prof. Dr. Michael Griebel

INSTITUT FÜR NUMERISCHE SIMULATION

MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT DER  
RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN



## Notation

Symbol	Bedeutung
$\mathbf{A}, \mathbf{B}$	Matrizen
$\mathbf{a}, \mathbf{b}, \dots, \mathbf{z}$	Vektoren
$a, b, \dots, z$	Skalare
$\mathbf{x}', \mathbf{X}'$	Der transponierte Vektor von $\mathbf{x}$ und die transponierte Matrix von $\mathbf{X}$
$\mathbf{I}d_n$	Einheitsmatrix der Größe $n \times n$
$\mathbf{I}d_{n \times m}$	Einheitsmatrix der Größe $n \times m$
$\Sigma$	Kovarianzmatrix
$\Lambda$	Diagonalmatrix
$\mathbf{L}$	Untere Dreiecks-Matrix einer Cholesky-Zerlung
$ \Sigma $	Determinante der Matrix $\Sigma$
$\ \mathbf{x}\ _1$	Spaltensummennorm von $\mathbf{x}$
$X, Y, Z$	Zufallsvariablen
$\mathbb{E}[X]$	Erwartungswert der Zufallsvariablen $X$
$\mathbb{V}[X]$	Varianz der Zufallsvariablen $X$
$Cov(X, Y)$	Kovarianz der Zufallsvariablen $X$ und $Y$
$\mathcal{N}(\mu, \sigma^2)$	Univariate Normalverteilung mit Erwartungswert $\mu$ und Varianz $\sigma^2$
$\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$	Multivariate Normalverteilung der Dimension $n$ mit Erwartungswerten $\boldsymbol{\mu}$ und Kovarianzmatrix $\Sigma$
$\mathcal{U}_d(\Omega)$	Die $d$ -dimensionale Gleichverteilung auf dem Gebiet $\Omega$
$\phi(x)$	Dichtefunktion der standardisierten Normalverteilung $\mathcal{N}(0, 1)$
$\boldsymbol{\phi}(x)$	Dichtefunktion von $n$ unabhängig und $\mathcal{N}(0, 1)$ verteilten Zufallsvariablen
$\Phi(x)$	Verteilungsfunktion der standardisierten Normalverteilung $\mathcal{N}(0, 1)$
$\mathbf{1}(x)$	Indikatorfunktion
$id(x)$	Identitätsfunktion
$\Gamma(x)$	Wert der Gammafunktion an der Stelle $x$
$Df(\mathbf{x})$	Differential der Funktion $f$ an der Stelle $\mathbf{x}$
$Q_l^{(d)} f$	Anwendung der Quadraturregel $Q$ mit Level $l$ auf die $d$ -dimensionale Funktion $f$
$\mathcal{C}^\infty$	Menge der unendlich oft differenzierbaren Funktionen
$L^2[a, b]$	Menge der zweifach Lebesgues-integrierbaren Funktion
$(f, g)_w$	Skalarprodukt auf dem Raum $L^2[a, b]$ mit Gewichtsfunktion $w(x)$



# Inhaltsverzeichnis

<b>Notation</b>	<b>i</b>
<b>Inhaltsverzeichnis</b>	<b>iv</b>
<b>1. Einleitung</b>	<b>1</b>
<b>2. Ökonometrische Grundlagen</b>	<b>7</b>
2.1. Motivation . . . . .	7
2.1.1. Das Okunsche Gesetz . . . . .	7
2.2. Lineare Regressionsmodelle . . . . .	8
2.2.1. Modellannahmen . . . . .	8
2.2.2. Parameterschätzung . . . . .	9
2.2.3. Heteroskedastizität und Autokorrelation . . . . .	10
2.3. Generalisierte lineare Modelle . . . . .	10
2.3.1. Kontinuierliche Variablen . . . . .	11
2.3.2. Häufigkeitsdaten . . . . .	12
2.4. Binary-Choice-Modelle . . . . .	12
2.4.1. Motivation und Modellaufbau . . . . .	12
2.4.2. Eigenschaften des Modells und Zusammenhang zur allg. Definition . . . . .	13
2.4.3. Modelle mit latenten Variablen . . . . .	14
2.4.4. Random Utility Modelle . . . . .	15
2.4.5. Parameterschätzung . . . . .	15
2.4.6. Beispiele für Verteilungsfunktionen . . . . .	16
2.5. Modelle für Panel-Daten . . . . .	17
2.6. Probit-Modelle . . . . .	19
2.6.1. Panel Probit-Modell . . . . .	19
2.6.2. Multivariate Probit-Modelle . . . . .	21
2.6.3. Multinomiale Probit-Modelle . . . . .	22
2.6.4. Ordered multivariate Probit-Modelle . . . . .	23
2.7. Logit-Modelle . . . . .	24
2.7.1. Multinomiale Logit-Modelle . . . . .	24
2.7.2. Mixed Logit-Modelle . . . . .	25
<b>3. Hochdimensionale Integration</b>	<b>29</b>
3.1. Monte Carlo . . . . .	29
3.2. Quasi-Monte Carlo . . . . .	30
3.3. Produktansatz . . . . .	30
3.3.1. Eindimensionale Quadraturverfahren . . . . .	31
3.4. Dünne Gitter . . . . .	33

<b>4. Transformationstechniken</b>	<b>35</b>
4.1. Cholesky-Zerlegung . . . . .	35
4.2. Deáks Ansatz . . . . .	36
4.3. Schervishs Methode . . . . .	37
4.4. Genz-Algorithmus/GHK-Simulator . . . . .	38
4.4.1. Variablen-Umordnungstechnik von Gibson . . . . .	40
4.4.2. Variablen-Umordnungstechnik von Genz . . . . .	41
4.5. PA-Simulator . . . . .	41
4.6. Effizientes hochdimensionales Importance Sampling . . . . .	42
4.7. Weitere Techniken . . . . .	44
<b>5. Numerische Resultate</b>	<b>45</b>
5.1. Performance-Vergleiche für ein Probit-Modell . . . . .	45
5.1.1. Anpassung der Daten . . . . .	45
5.1.2. Autoregressive Korrelationsstruktur . . . . .	46
5.1.3. Numerische Ergebnisse . . . . .	48
5.2. Anwendung auf ein Mixed Logit-Modell . . . . .	53
5.3. Panel Probit-Analyse . . . . .	54
<b>6. Zusammenfassung und Ausblick</b>	<b>57</b>
<b>A. Anhang</b>	<b>59</b>
A.1. Maximum-Likelihood-Schätzung . . . . .	59
A.2. Transformationssatz . . . . .	59
A.3. Szántais Ansatz . . . . .	60
<b>Literaturverzeichnis</b>	<b>65</b>

## 1. Einleitung

Die Ökonometrie ist ein wichtiges Teilgebiet der Wirtschaftswissenschaften, das die Modelle der Wirtschaftstheorie empirisch überprüft und versucht die auftretenden ökonomischen Phänomene zu quantifizieren. Hierzu werden mathematische Methoden mit statistischen Verfahren kombiniert. Auch in anderen empirischen Sozialwissenschaften finden ökonometrische Methoden vermehrt Anwendung.

Während das klassische Werkzeug der Ökonometrie, die lineare Regressionsanalyse, nicht an Bedeutung verloren hat, erweitern moderne computerorientierte Verfahren die existierenden Methoden. Letztere erlauben die Analyse auch von extrem großen Datensätzen und bieten empirischen Forschern eine größere Freiheit bei der Modellierung von Zusammenhängen und Abhängigkeiten.

Der Analysegegenstand vieler wirtschaftstheoretischer Modelle, die durch ökonometrische Methoden überprüft werden, sind menschliche Entscheidungsprobleme. Während einige dieser Entscheidungen einen kontinuierlichen Charakter haben, d.h. eine Entscheidung wird über die bestimmte Menge getroffen, sind andere wichtige Entscheidungsprobleme von diskreter Natur. Das bedeutet, dass Entscheidungen zwischen einer festen Anzahl von genau definierten Alternativen getroffen werden.

Beispiele für Entscheidungsprobleme diskreter Art sind die Wahl zwischen mehreren politischen Parteien, Kaufentscheidungen zwischen einer endlichen Anzahl von Produkten oder auch die Entscheidung über einen Arztbesuch in der Gesundheitsökonomie.

Die geeigneten Modelle zur genauen Untersuchung solcher diskreten Entscheidungen und ihrer Einflussfaktoren sind die *Discrete Choice Modelle* der Ökonometrie. Aufgrund der Vielzahl von Anwendungen, der immer flexibleren Modellierungsmöglichkeiten durch theoretische Fortschritte und aufgrund der verbesserten Verfügbarkeit geeigneter Datensätze haben sie enorm an Bedeutung gewonnen.

Die Discrete Choice Modelle ermöglichen die Analyse von menschlichen Entscheidungen und können damit als ein Hilfsmittel zum besseren Verständnis menschlichen Verhaltens verstanden werden. Ihre Bedeutung ist auch erkennbar an der Verleihung des von der schwedischen Reichsbank in Erinnerung an Alfred Nobel gestifteten Preises für Wirtschaftswissenschaften (im allgemeinen Sprachgebrauch „Wirtschaftsnobelpreis“) an Daniel McFadden für „die Entwicklung von Theorie und Methoden zur Analyse diskreter Entscheidungen“ im Jahr 2000.

Während die Analyse von einzelnen binären Entscheidungen bereits durch zufriedenstellende Methoden möglich ist, ist dieses für multikategoriale Entscheidungen nur begrenzt oder nur mit sehr spezifischen Modellannahmen möglich. Multikategoriale Entscheidungen sind dabei alle Arten von Entscheidungen, die über eine binäre Entscheidung hinausgehen, wie z.B. Modelle zur Analyse von mehr als zwei Wahlmöglichkeiten oder Modelle zur Analyse einer binären Entscheidung an mehreren Zeitpunkten.

## Problemstellung

Die zur Analyse von diskreten Entscheidungen verwendeten Modelle lassen sich durch die Gleichung

$$\mathbb{P}[\text{Wahl von Alternative } i] = F(\text{Parameter, Daten}) \quad (1.1)$$

darstellen. Es wird also die Wahrscheinlichkeit für die Wahl einer bestimmten Alternative in Abhängigkeit von Parametern und ihren Einflussfaktoren, den unabhängigen Variablen, untersucht. Die Verknüpfung der Entscheidungswahrscheinlichkeit mit den Parametern und Daten erfolgt über eine Funktion  $F$ , die *Responsefunktion*.

Während in einigen Modellformen und Zusammenhängen die Wahl von  $F(x)$  als eine beliebige, monotone Funktion  $F : \mathbb{R}^d \rightarrow [0, 1]$  erfolgen kann (siehe [FT01]), wird sie in anderen Modellen durch die Annahme an die stochastischen Störungen des Modells determiniert und ist dann durch die Verteilungsfunktion der den Störungen zugrundeliegenden Zufallsvariablen gegeben. Wichtige Modelle dieser Art sind Untersuchungen mit *latenten* Variablen. Diese analysieren einen Zusammenhang mit abhängigen Variablen, die grundsätzlich von kontinuierlicher Art, jedoch nicht beobachtbar sind. Es kann lediglich eine Variable beobachtet werden, die eine Indikation über die quantitative Höhe der eigentlichen Variablen gibt.

Ein einfaches Beispiel dieser Art ist die Untersuchung der Bedeutung (des Nutzens) des Gesundheitszustandes für verschiedene Individuen in Abhängigkeit des Alters, des Einkommens und anderer bestimmender Größen. Eine Indikation über die Bedeutung des Nutzens von Gesundheit für ein Individuum könnte die binäre Entscheidung für oder wider einen Arztbesuch bilden.

Neben den beobachteten Daten hängt das Modell in Gleichung (1.1) von Parametern ab, die quantifizieren wie sich die Entscheidungswahrscheinlichkeit für eine Alternativ ändert, falls sich die gegebenen Daten verändern. Die Schätzung dieser Parameter ist eine zentrale Aufgabe von ökonomischen Verfahren und wird über die Maximum-Likelihood-Methode durchgeführt. Die hierzu nötige Likelihood-Funktion muss zunächst aus der rechten Seite von Gleichung (1.1) mit den gegebenen Daten berechnet werden und sollte möglichst exakt sein, damit ihre Maximierung bestmöglich erfolgen kann. Insbesondere ist für eine einmalige Auswertung der Likelihood-Funktion eine mehrmalige Auswertung von  $F$  erforderlich.

Da  $F$  in allen relevanten Fällen durch die Verteilungsfunktion einer Zufallsvariablen gegeben ist, kann die Auswertung von  $F$  auch als ein Integrationsproblem der Form

$$F(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f(\mathbf{x}) d\mathbf{x} \quad (1.2)$$

mit einer Dichtefunktion  $f$  interpretiert werden. Je nach Zufallsvariable kann eine geschlossene Lösung für  $F$  existieren oder nicht.

Während für binäre Entscheidungen die Responsefunktion  $F$  (und damit das Integrationsproblem (1.2)) eindimensional ist und daher hinreichend gute Approximationen der rechten Seite in Gleichung (1.2) (oder sogar die analytische Bestimmung dieser) möglich sind, ist die Funktion  $F$  für multikategoriale Entscheidungen mehrdimensional. Die Dimension  $d$  des Problems kann dabei durch die Anzahl der Wahlmöglichkeiten, die Anzahl der untersuchten Zeitpunkte oder auch die Anzahl der unabhängigen Variablen determiniert sein und beliebig groß werden.

---

Der kritische Punkt dieser Methode ist die Auswertung der Funktion  $F$  für multikategoriale Modelle, falls keine geschlossene Lösung für  $F$  existiert, da dann eine numerische hochdimensionale Integration erforderlich ist. Dieses ist für die multikategorialen Erweiterungen der beiden wichtigsten Modelltypen, die Probit- und die Logit-Modelle, der Fall. Erstere spezifizieren  $F$  als Verteilungsfunktion der multivariaten Normalverteilung und stehen im Hauptfokus dieser Arbeit.

Die Bestimmung des Integrals in Gleichung (1.2) kann durch Quadraturformeln der Form

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} \approx Qf := \sum_{i=1}^N \omega_i f(\mathbf{x}^{(i)})$$

erfolgen. Die numerische Bestimmung dieser Integrale ist mit herkömmlichen Produkt-Quadraturverfahren schwer möglich da deren Konvergenzrate exponentiell mit der Dimension des Problems fällt. Dieses Phänomen wird auch als *Fluch der Dimension* (siehe [Bel61]) bezeichnet. Für eine ausführliche Darstellung dieses Phänomens und geeigneter Lösungsmöglichkeiten siehe auch [Gri06].

Ein weiteres Problem ist das unbeschränkte Gebiet des Integrationsproblems (1.2). In dieser Arbeit zeigen wir Möglichkeiten auf beide Problem zu lösen und eine gute Approximation der rechten Seite von (1.1) zu bestimmen.

## Lösungsansätze

Zur Lösung des uneigentlichen Integrationsproblems, der Auswertung einer multivariaten Normalverteilung, ist die Transformation des ursprünglichen Problems von großer Bedeutung. Hierzu sind verschiedene Ansätze entwickelt worden, welche die Transformation des Integrationsproblems auf ein anderes Integrationsgebiet vereint und die Konvergenz der numerischen Methoden dadurch verbessert.

Wichtige Beispiele sind die Transformation auf ein sphärisches Koordinatensystem von Deák in [Dea80] und die Transformation auf den Einheitswürfel  $(0,1)^{d-1}$ . Letztere ist unabhängig voneinander von Genz in [Gen92] in der Statistik und in der Ökonometrie durch die Gruppe Geweke und Hajivassiliou (siehe [Gew89] und [BSH93]) sowie Keane in [Kea94] entwickelt worden und hat sich als besonders erfolgreich erwiesen. Eine Weiterentwicklung der Methode ist durch eine Umordnungstechnik, die eine Ordnung entlang der Wichtigkeit der Dimension einführt, von Gibson et al. in [GGE94] beschrieben worden

Die diskutierte Transformationstechnik ist ursprünglich für die Auswertung des Integrals mit Zufallszahlen, d.h. einem Monte Carlo Verfahren, entwickelt worden. In dieser Arbeit diskutieren wir sowohl die Verwendung von Quasi-Monte Carlo Verfahren als auch die Verwendung von dünnen Gittern zur Lösung des transformierten Integrationsproblems. Quasi-Monte Carlo Verfahren verwenden als Auswertungspunkte Folgeglieder von Niederdiskrepanz-Folgen anstatt gleichverteilten Zufallszahlen und können so die Konvergenz gegenüber (Pseudo-)Zufallszahlen erhöhen. Dünne Gitter können die Glattheit des Integranden ausnutzen und auf diese Weise die Konvergenz verbessern.

Die Dünngitter-Integration kombiniert aus gelevelten Tensorprodukten eindimensionale Quadraturregeln zu einer  $d$ -dimensionalen Quadraturregel. Als eindimensionale Basisregel verwenden

wir dabei die geschachtelten Clenshaw-Curtis-Regeln und generalisierte Gauß-Quadraturregeln. Letztere sind besonders für das Integrationsproblem geeignet, da sie die bei der Transformation nach Genz entstehenden Randsingularitäten berücksichtigen können.

Es gelingt uns zu zeigen, dass die dünnen Gitter für das Integrationsproblem geeignet sind. Im Speziellen führt die Genz-Transformation mit den dünnen Gittern auf Basis der generalisierten Gauß-Quadraturregeln zu einer deutlich verbesserten Approximation des Integrals.

Auch die in dieser Arbeit erstmals im Kontext der dünnen Gitter betrachtete Umordnungstechnik von Gibson et al. führt zu einer Verbesserung aller betrachteten Integrationsverfahren. Die Verbesserung der Integration durch die Umordnungstechnik ist bei Verwendung der Dünngitter-Integration besonders groß. Zwar führt sie nicht zu einer Erhöhung der Konvergenzrate des Verfahrens, jedoch kann der Einfluss auf die Konstante erheblich sein. Dadurch kann die benötigte Anzahl von Auswertungspunkten für eine bestimmte Genauigkeit deutlich gesenkt werden.

Insgesamt kann durch die Kombination der Genz-Transformation, der Umordnungstechnik nach Gibson et al. und den dünnen Gittern mit den generalisierten Gauß-Quadraturregeln ein zufriedenstellendes Verfahren für die Auswertung der multivariaten Normalverteilung für niedrige und mittlere Dimensionen aufgezeigt werden.

Auf diese Weise ist eine exaktere Auswertung der Funktion  $F$  in den Discrete-Choice-Modellen und damit eine bessere Approximation der Likelihood-Funktion möglich.

### Eigene Beiträge

- Herausarbeitung des Zusammenhangs zwischen Discrete-Choice-Modellen und generalisierten linearen Modellen, sowie Identifizierung hochdimensionaler Probleme im Kontext der Discrete-Choice-Modelle.
- Vergleich existierender Methoden zur Berechnung der multivariaten Normalverteilung und Entwicklung einer neuen verbesserten Methode durch die Kombination von Transformations- und Umordnungstechniken mit dünnen Gittern.
- Schätzung eines Modells mit realen Daten.

### Aufbau der Arbeit

In Kapitel 2 geben wir zunächst eine kurze Einführung in die grundlegenden linearen Regressionsmodelle der Ökonometrie und entwickeln darauf aufbauend die Struktur der generalisierten linearen Modelle und insbesondere der hierin enthaltenen Discrete-Choice-Modelle. Der Fokus liegt hierbei auf den entstehenden hochdimensionalen Fällen dieser Modelle sowie auf den für die Schätzung der Parameter benötigten Größen. Des Weiteren werden konkrete ökonomische Anwendungen aufgezeigt.

In Kapitel 3 stellen wir verschiedene Verfahren für die hochdimensionale Integration dar und beschreiben ihre Eigenschaften.

In Kapitel 4 erläutern wir unterschiedliche Ansätze zur Auswertung der multivariaten Normalverteilung, die in einer Vielzahl von Fällen als Integrationsproblem im Kontext der Discrete-Choice-Modelle auftritt. Wir betrachten insbesondere verschiedene Verfahren zur Transforma-

---

tion des Integrationsproblems, so dass dieses zur numerischen Integration mit dünnen Gittern besser geeignet ist.

Das Kapitel 5 dient den numerischen Experimenten. Wir untersuchen die Eigenschaften unterschiedlicher Integrations- und Transformationsverfahren für multivariate Normalverteilungen, die sich als Integrationsproblem der Probit-Modelle ergeben und erweitern unsere Analyse dann auf ein Integrationsproblem, das sich im Rahmen der multikategorialen Logit-Modelle ergibt. Während wir zur Analyse generierte Daten verwenden, führen wir abschließend eine Maximierung einer approximierten Likelihood-Funktion mit einem realen Datensatz aus der Gesundheitsökonomie durch und belegen damit die praktische Relevanz der von uns entwickelten Methode.

Kapitel 6 fasst die Ergebnisse dieser Arbeit zusammen und gibt einen Ausblick auf mögliche Erweiterungen, sowohl in Bezug auf die Anwendbarkeit der diskutierten Verfahren in anderen Kontexten als auch in Bezug auf die Weiterentwicklung des Verfahrens selbst.

## Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Zunächst möchte ich mich bei Prof. Dr. Michael Griebel für die Vergabe eines sehr interessanten Themas, das meine Studienfächer perfekt verknüpft, sowie für die zahlreichen Anregungen und Hilfestellungen bei Problemen bedanken. Des Weiteren danke ich Prof. Dr. Jochen Garcke für die Übernahme des Zweitgutachtens.

Meinem Betreuer Jens Oettershagen, welcher durch seine vielen Anregungen, Erläuterungen, Korrekturen, seine Geduld bei Fragen und sein Verständnis die Arbeit in diesem Ausmaß erst ermöglicht hat, gilt mein besonderer Dank.

Bei allen Mitarbeitern des Instituts für Numerische Simulation bedanke ich mich für Hilfestellungen, Diskussionen und eine sehr angenehme Arbeitsatmosphäre, insbesondere bei Viktoria Gleiser und Nadine Pastuszka.

Schließlich danke ich meinen Eltern und meinen Freunden für ihre langjährige Unterstützung, ihr Verständnis und besonders für ihre Geduld.



## 2. Ökonometrische Grundlagen

In diesem Kapitel erläutern wir die wichtigen ökonometrischen Grundlagen, die zum Verständnis des in dieser Arbeit betrachteten Anwendungszusammenhangs nötig sind.

Nach einer kurzen Motivation definieren wir die grundlegenden linearen Regressionsmodelle der Ökonometrie und erläutern die für diese Art von Modellen übliche Schätzmethodik. Darauf aufbauend erweitern wir das Modell zu den generalisierten linearen Modellen und definieren die Komponenten dieser umfassenden Modelle.

Wir führen Beispiele für solche Modelle an und fokussieren uns dabei im Besonderen auf eine wichtige Klasse der generalisierten linearen Modelle, die Discrete-Choice-Modelle.

Im letzten Teil dieses Kapitels gehen wir dann in besonderem Maße auf die hochdimensionalen Fälle der beiden wichtigsten Modelle, die Probit- und Logit-Modelle, ein.

### 2.1. Motivation

In der Erstausgabe der Zeitschrift *Econometrica* erläutert die Econometric Society, dass es das Hauptziel der Ökonometrie [als Teilgebiet der Volkswirtschaftslehre] sei eine Verbindung von theoretisch-quantitativen und empirischen-quantitativen Ansätzen zu ökonomischen Problemen herzustellen. Insbesondere sei es die Verbindung von Sichtweisen aus der Statistik, der ökonomischen Theorie und der Mathematik, die die Ökonometrie ausmache (siehe [Fri33]).

Eines der Kerngebiete der Ökonometrie ist die Schätzung von Parametern in ökonomischen Modellen, die Zusammenhänge zwischen einzelnen Größen wie z.B. Inflationsrate, Bruttoinlandsprodukt und Arbeitslosigkeit abbilden. Ein prominentes Beispiel für ein solches, ökonomisches Modell ist das Okunsche Gesetz.

#### 2.1.1. Das Okunsche Gesetz

In [Oku62] untersucht Okun den Zusammenhang zwischen Veränderungen der Arbeitslosenrate und Veränderungen des realen Bruttoinlandsprodukts (BIP). Eine seiner Modellierungen stellt diesen Zusammenhang über die lineare Beziehung

$$y = a + bx$$

her, wobei  $x$  die quartalsweise prozentuale Veränderung des BIPs und  $y$  die quartalsweise prozentuale Veränderung der Arbeitslosenrate enthält. Mit Daten für die Jahre 1947-1960 für die US-amerikanische Volkswirtschaft und der in Abschnitt 2.2.2 erläuterten Methode kommt er auf den Zusammenhang

$$y = 0.3 - 0.3x.$$

Dies bedeutet, dass eine Steigerung des BIPs um 1% mit einer Verringerung der Arbeitslosenrate um 0.3% einhergeht.

## 2.2. Lineare Regressionsmodelle

Das Okunsche Gesetz ist ein Beispiel für ein lineares Regressionsmodell. Ein solches Modell untersucht den Zusammenhang zwischen einer **abhängigen, erklärten** oder **endogenen** Variablen (auch **Regressand**)  $y$  und einer oder mehreren **unabhängigen, erklärenden** oder **exogenen** Variablen (auch **Regressoren**)  $x_i$  durch eine lineare Gleichung. Dieses lässt sich mathematisch als

$$\begin{aligned} y &= f(x_1, \dots, x_n) + \epsilon \\ &= x_1\beta_1 + \dots + x_K\beta_K + \epsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \epsilon \end{aligned} \tag{2.1}$$

mit  $\mathbf{x} = (x_1, \dots, x_K)'$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)' \in \mathbb{R}^{K \times 1}$  darstellen.

Der zufällige Störterm  $\epsilon$  trägt hierbei dem Umstand Rechnung, dass es i.d.R. nicht gelingen wird den realen ökonomischen Zusammenhang exakt zu modellieren und dass es zu Messfehlern bei modellrelevanten Daten kommen kann (siehe [Gre08b] Kap. 2.2.).

Zur Spezifikation der Modellannahmen seien die  $n$  Beobachtungen einer Variablen  $x_i$  in einem Vektor  $\mathbf{x}_i \in \mathbb{R}^{n \times 1}$  und diese Vektoren  $\mathbf{x}_i$  in einer Matrix  $\mathbf{X} \in \mathbb{R}^{n \times K}$  zusammengefasst. Ferner seien die  $n$  Beobachtungen der Variablen  $y$  und die Störterme  $\epsilon$  in Vektoren  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  bzw.  $\boldsymbol{\epsilon} \in \mathbb{R}^{n \times 1}$  zusammengefasst.

In diesem Modell ist es auch möglich eine Konstante in der Gleichung (2.1) zu modellieren, was dazu führt, dass eine entsprechende Spalte von  $\mathbf{X}$  durch den Vektor  $\mathbf{I}\mathbf{d}_{n \times 1} = (1, \dots, 1)' \in \mathbb{R}^{n \times 1}$  gegeben ist.

### 2.2.1. Modellannahmen

Um das Modell exakt zu spezifizieren werden die folgenden Annahmen ([Gre08b] Tab. 2.1) an das Modell gestellt:

1. **Linearität:** Das Modell bildet einen linearen Zusammenhang zwischen  $y$  und  $x_1, \dots, x_K$  wie in Gleichung (2.1) ab. Da das für jeden der  $n$  Datenpunkte gelten soll, lässt sich dieses darstellen als  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ .
2. **Voller Rang:** Es gibt keine exakte lineare Abhängigkeit zwischen den erklärenden Variablen, d.h.  $\text{rang}(\mathbf{X}) = K$ .
3. **Exogene erklärende Variablen:** Die erklärenden Variablen  $x_i$  enthalten keine Informationen über den Störterm  $\epsilon$ , d.h.  $\mathbb{E}[\boldsymbol{\epsilon}|\mathbf{X}] = \mathbf{0}$ .
4. **Homoskedastizität und fehlende Autokorrelation:** Jeder Fehlerterm  $\epsilon_i$  hat die konstante Varianz  $\sigma^2$  (Homoskedastizität) und ist unkorreliert mit allen anderen  $\epsilon_j$  (Freiheit von Autokorrelation). Dieses lässt sich beschreiben als  $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \sigma^2\mathbf{I}\mathbf{d}_n$ .
5. **Daten:** Die  $n$  Datensätze für die erklärenden Variablen können deterministisch oder Ausprägungen einer Zufallsvariablen sein. In Matrixform:  $\mathbf{X}$  sei deterministisch oder zufällig.
6. **Normalverteilung:** Die Fehlerterme  $\epsilon_i$  sind normalverteilt, d.h.  $\boldsymbol{\epsilon}|\mathbf{X} \sim \mathcal{N}_n(0, \sigma^2\mathbf{I}\mathbf{d}_n)$ .

### 2.2.2. Parameterschätzung

#### Kleinste-Quadrate-Schätzung

Ein möglicher Ansatz zur Bestimmung der Modellparameter  $\beta$  ist die Minimierung der entstehenden Fehlerquadrate. Dazu seien  $\beta$  und  $\epsilon$ , wie bisher, die unbekannt Parameter bzw. Störterme des Modelles. Des Weiteren enthalte  $\mathbf{b}$  Schätzungen für  $\beta$  und  $\mathbf{e}$  entsprechend für  $\epsilon$ . Es sei  $\hat{\mathbf{y}}$  eine Schätzung für die Werte von  $\mathbf{y}$ , die bei Benutzung der geschätzten Parameter  $\mathbf{b}$  entsteht. Dieses lässt sich zusammenfassen als

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\mathbf{b} \\ \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}}.\end{aligned}$$

Offenbar enthält der Vektor  $\mathbf{e}$  die Abstände, auch Residuen, zwischen den tatsächlich realisierten Werten von  $\mathbf{y}$  und den prognostizierten  $\hat{\mathbf{y}}$ . Als ein sinnvolles Güte-Kriterium für  $\mathbf{b}$  erscheint die Minimierung dieser Abstände. Um dieses Vorzeichen unabhängig zu gestalten, werden die Fehlerquadrate minimiert:

$$\begin{aligned}S(\mathbf{b}) &:= \sum_{i=1}^n e_i^2 = \mathbf{e}'\mathbf{e} \\ &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\ &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}.\end{aligned}$$

Die notwendige Bedingung für das Minimum ist

$$\begin{aligned}\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = 0. \\ &\Leftrightarrow \mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.\end{aligned}\tag{2.2}$$

Es ist zu erkennen, dass die Bedingung für das Minimum der von linearen Ausgleichsproblemen bekannten Normalgleichung (siehe [Sto94] Kap. 4.8.1) entspricht. De facto kann die Parameterschätzung im linearen Regressionsmodell als ein solches Ausgleichsproblem interpretiert werden.

Aus (2.2) ergibt sich als Schätzer für die Parameter

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

#### Maximum-Likelihood-Schätzung

Lineare Regressionsmodelle können auch mit Hilfe der Maximum-Likelihood-Methode geschätzt werden. Die Methode ist im Anhang A.1 dargestellt. Insbesondere lässt sich zeigen, dass diese Methode im Falle von linearen Regressionsmodellen zum gleichen Schätzer für  $\beta$  führt (siehe [Gre08b] Kap. 14.2.1).

### 2.2.3. Heteroskedastizität und Autokorrelation

Eine Erweiterung des linearen Regressionsmodells ist das **verallgemeinerte** lineare Regressionsmodell, das die 4. Annahme des ursprünglichen Modells aufhebt und sowohl Autokorrelation als auch Heteroskedastizität (als Gegenstück zu Homoskedastizität) zulässt. Das ist gleichbedeutend mit der Annahme einer allgemeinen Kovarianz-Matrix für die Störterme  $\epsilon$ ,

$$\mathbb{E}[\epsilon\epsilon'] = \Sigma$$

mit einer positiv-definiten Matrix  $\Sigma$ .

Auch für die verallgemeinerten linearen Modelle gibt es verschiedene Herangehensweisen an die Parameterschätzung, die z.B. auch davon abhängen, ob  $\Sigma$  unbekannt ist oder nicht. Für eine ausführliche Darstellung siehe [Gre08b] Kap. 8.

## 2.3. Generalisierte lineare Modelle

Nach der ausführlichen Darstellung der linearen Regressionsmodelle wollen wir in diesem Abschnitt die soeben diskutierten Modelle auf die „Generalized Linear Models“, die generalisierten linearen Modelle erweitern. Sie sind ein sehr weitgehendes Framework aus der Statistik, das die Modellierung von unterschiedlichsten Zusammenhängen und Datentypen erlaubt. Während das lineare Regressionsmodell offenbar nur die Modellierung von Zusammenhängen erlaubt, in denen die abhängige Variable kontinuierlich ist, ist durch die generalisierten linearen Modelle z.B. auch die Modellierung von abhängigen Variablen mit binärer Struktur oder die Modellierung von abhängigen Variablen, die eine bestimmte Anzahl beschreiben, möglich.

Zur Verdeutlichung der allgemeinen Definition stellen wir zunächst basierend auf [MN89] Kap.2.2 fest, dass sich das lineare Regressionsmodell auch wie folgt beschreiben lässt:

Es sei  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  mit unabhängigen und identisch normalverteilten Einträgen  $y_i$  mit  $\mathbb{V}[y_i] = \sigma^2$  für  $i = 1, \dots, n$  gegeben und es gelte

$$\mathbb{E}[\mathbf{y}] = \boldsymbol{\mu} \tag{2.3}$$

$$\text{mit } \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \tag{2.4}$$

für Beobachtungen  $\mathbf{X} \in \mathbb{R}^{n \times K}$  und Parameter  $\boldsymbol{\beta} \in \mathbb{R}^{K \times 1}$ .

Ausgehend von dieser Beobachtung folgen wir der dreigliedrigen Definition für generalisierte lineare Modelle von McCullagh et al. in [MN89] Kap. 2.2.1.:

1. **Zufallskomponente:** Es sei  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  ein Vektor von unabhängigen Zufallsvariablen (die abhängigen Variablen des Modells) aus der Familie der exponentiellen Verteilungen mit konstanter Varianz  $\sigma^2$  und  $\mathbb{E}[\mathbf{y}] = \boldsymbol{\mu}$ .

Die Dichte einer solchen Zufallsvariablen aus  $\mathbf{y}$  ist in diesen Modellen durch

$$f(y_i|\theta_i, \phi, \omega_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi}\omega_i + c(y_i, \phi, \omega_i)\right)$$

mit

- a)  $\theta_i$  als „natürlichem“ Parameter,

- b)  $\phi$  als Skalierungsparameter,
- c)  $b(x)$  und  $c(x)$  als spezielle Funktionen der Verteilung und
- d)  $\omega_i$  als Gewicht abhängig vom Datentyp

gegeben (siehe [FT01] Kap. 2.1.2). Für eine Darstellung der Eigenschaften dieses Typs von Zufallsvariablen verweisen wir auf den Anhang von [FT01].

2. **Systematische Komponente:** Es sei  $\boldsymbol{\eta} \in \mathbb{R}^{n \times 1}$  ein linearer Prädiktor, der sich aus den abhängigen Variablen des Modells durch

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

mit unbekanntem Parameter  $\boldsymbol{\beta} \in \mathbb{R}^{K \times 1}$  ergibt.

3. **Linkfunktion:** Die Zufallskomponente und die systematische Komponente seien für  $i = 1, \dots, n$  durch eine auf dem Wertebereich der Zufallsvariablen  $y_i$  monotone und differenzierbare Linkfunktion  $G(\mu)$  bzw. ihrem Inversen, der Responsefunktion  $H(\eta) = G^{-1}(\eta)$ , mit

$$\eta_i = G(\mu_i) \Leftrightarrow \mu_i = H(\eta_i)$$

verbunden.

Das lineare Modell aus den Gleichungen (2.3) und (2.4) ist in dieser Definition enthalten, indem die Einträge von  $\mathbf{y}$  als normalverteilt und  $g(x)$  als die Identitätsfunktion spezifiziert wird.

Daraus wird ersichtlich, dass die beiden wesentlichen Erweiterungen in den generalisierten linearen Modellen (gegenüber dem linearen Regressionsmodell) einerseits durch die Verallgemeinerung der Verteilungsannahme an die Einträge der abhängigen Variablen  $\mathbf{y}$  und andererseits durch die Möglichkeit verschiedener Linkfunktionen gegeben sind. Die erste Annahme ist insbesondere auch gleichbedeutend mit der Annahme einer von der Normalverteilung verschiedenen Verteilung der Störterme  $\boldsymbol{\epsilon}$  im Fall des linearen Regressionsmodells.

Während die obige Definition als Response- bzw. Linkfunktion alle Funktionen einer Klasse mit bestimmten Eigenschaften zulässt, induziert jede Zufallsvariable eine kanonische Linkfunktion. Dieses ist die Funktion, die den natürlichen Parameter  $\theta$  der Dichtefunktion der zugrundeliegenden Zufallsvariablen mit dem linearen Prädiktor  $\eta$  verbindet:

$$G(\mu) = \eta = \theta(\mu).$$

In Folgenden geben wir Beispiele für verschiedene Datentypen der abhängigen Variablen, möglichen Zufallsvariablen zur Modellierung solcher Daten und passende Responsefunktionen an. Wir verzichten dabei auf die explizite Angabe der Dichtefunktion für jede Zufallsvariable und geben jeweils nur eine kurze Intuition für den Datentyp. Die Dichtefunktionen können leicht aus der Tabelle 2.1. in [FT01] entnommen werden, während [MN89] Kap. 2.2.4 einen Überblick über die kanonischen Linkfunktionen für die jeweiligen Zufallsvariablen gibt.

### 2.3.1. Kontinuierliche Variablen

Falls wir einen Anwendungszusammenhang betrachten, in dem die abhängigen Variablen jeden Wert in  $\mathbb{R}$  annehmen dürfen (z.B. das motivierende Beispiel in Abschnitt 2.1.1), bietet sich als

Verteilungsannahme für die stochastische Komponente des Modells die Normalverteilung an. Neben der Identitätsfunktion als kanonische Link- und Responsefunktion nennen Fahrmeier et al. in [FT01] Kap. 2.1.3 z.B. noch die Responsefunktion

$$H(\eta) = \log(\eta), \text{ und } H(\eta) = \exp(\eta)$$

für diese Verteilung.

In manchen statistischen Zusammenhängen kann die abhängige Variable zwar als kontinuierlich modelliert werden, allerdings nur auf dem Bereich  $(0, \infty)$ . Anwendungsbeispiele könnten die Lebenserwartung oder die Sonnenscheindauer eines Tages darstellen. Für diese Variablen bietet sich neben der Verwendung der Gammaverteilung mit der kanonischen Responsefunktion

$$h(\eta) = \frac{1}{\eta},$$

auch noch die Verwendung der inversen Normalverteilung an.

### 2.3.2. Häufigkeitsdaten

Weitere mögliche Anwendungszusammenhänge könnte die Untersuchung der Anzahl bestimmter Ereignisse in einem festgelegten Zeitraum sein. Beispiele für solche Häufigkeitsdaten, auch „Count Data“, sind die Anzahl von Geburten oder die Häufigkeit von Unfällen in einem bestimmten Zeitfenster. Diese Art von Daten wird typischerweise mit einer Poisson-Verteilung modelliert, deren Responsefunktion durch

$$h(\eta) = \exp(\eta)$$

gegeben ist.

## 2.4. Binary-Choice-Modelle

Im vorgehenden Abschnitt haben wir ein allgemeines Framework für „Generalized Linear Models“ eingeführt und verschiedene Beispiele aufgezeigt. Ein weiteres Beispiel für die generalisierten linearen Modelle sind Modelle, die binäre Daten untersuchen. Da auf diesem Datentyp (und seinen Erweiterungen) der Schwerpunkt dieser Arbeit liegt, stellen wir ihn ausführlich dar, leiten den Zusammenhang zur allgemeinen Definition in Abschnitt 2.3 ausführlich her und geben die typischen Kontexte aus der ökonomischen Theorie an, die zu diesen Modellen führt. Wir geben dabei insbesondere eine aus der ökonometrischen und nicht aus der statistischen Literatur motivierte Einführung.

### 2.4.1. Motivation und Modellaufbau

In vielen ökonomischen Anwendungen ist die zu untersuchende abhängige Variable als diskret zu betrachten. Zum Beispiel könnte der Zusammenhang zwischen Partizipation am Arbeitsmarkt (d.h. beschäftigt oder aktiv jobsuchend im Gegensatz zu weder beschäftigt noch jobsuchend) und verschiedenen Regressoren (z.B. Alter und Ausbildung) untersucht werden. Da hier nur

die Möglichkeit zwischen zwei verschiedenen Antworten (Partizipation oder nicht) besteht, wird auch von Binary Choice-Modellen gesprochen. Ein anderes Modellierungsbeispiel ist die Analyse des politischen Wahlverhaltens in einem Zweiparteiensystem.

Für die Modellspezifikation sei  $y$  der zu erklärende Regressand,  $\mathbf{x} \in \mathbb{R}^{K \times 1}$  ein Vektor mit den Werten für die  $K$  Regressoren, die die binäre Entscheidung erklären und  $\boldsymbol{\beta} \in \mathbb{R}^{K \times 1}$  ein Vektor mit Parametern, die den Einfluss der Veränderungen in  $\mathbf{x}$  erklären. Die Verbindung zwischen den Regressoren und der zu erklärenden Variable in der Entscheidung

$$\begin{aligned}\mathbb{P}(y = 1|\mathbf{x}) &= F(\mathbf{x}'\boldsymbol{\beta}) \\ \mathbb{P}(y = 0|\mathbf{x}) &= 1 - F(\mathbf{x}'\boldsymbol{\beta}).\end{aligned}\tag{2.5}$$

wird durch die Funktion  $F(x)$  spezifiziert. Die natürlichste Wahl für  $F(x)$  wäre die Identitätsfunktion  $id(x)$ , was (wieder) zu einem linearen Regressionsmodell führen würde. Jedoch ist bei dieser Wahl nicht sichergestellt, dass  $\mathbf{x}'\boldsymbol{\beta} \in [0, 1]$ , was aber notwendig ist, da  $F(\mathbf{x}'\boldsymbol{\beta})$  in (2.5) eine Wahrscheinlichkeit darstellt.

Wir stellen zusätzlich an das Modell die Bedingung, dass ein hoher numerischen Wert von  $\mathbf{x}'\boldsymbol{\beta}$  für gegebene Daten  $\mathbf{x}$  und gegebene Parameter  $\boldsymbol{\beta}$  zu einer hohen Wahrscheinlichkeit führt. Dieses bedeutet mathematisch, dass

$$\begin{aligned}\lim_{\mathbf{x}'\boldsymbol{\beta} \rightarrow +\infty} \mathbb{P}(y = 1|x) &= 1 \\ \lim_{\mathbf{x}'\boldsymbol{\beta} \rightarrow +\infty} \mathbb{P}(y = 0|x) &= 0\end{aligned}$$

gelten soll. Aus dieser Darstellung ist zu erkennen, dass jede Verteilungsfunktion einer reellwertigen Zufallsvariablen diese Anforderungen erfüllt. Verschiedene Binary-Choice-Modelle unterscheiden sich also in der Spezifikation dieser Verteilungsfunktion  $F(x)$ . Die laut Greene in [Gre08b] Kap. 23.3.1 gebräuchlichsten Modelle sind die in Abschnitt 2.4.6 dargestellten Probit- und Logit-Modelle.

### 2.4.2. Eigenschaften des Modells und Zusammenhang zur allg. Definition

Da die zu erklärende Variable  $y_i$  Bernoulli-verteilt mit Parameter  $G(\mathbf{x}_i'\boldsymbol{\beta})$  ist, ergibt sich für den Erwartungswert

$$\mathbb{E}[y_i|x_i] = F(\mathbf{x}_i'\boldsymbol{\beta})\tag{2.6}$$

und für die Varianz

$$\mathbb{V}[y_i] = F(\mathbf{x}_i'\boldsymbol{\beta})(1 - G(\mathbf{x}_i'\boldsymbol{\beta})).$$

Wenn wir Gleichung (2.6) in die Teile

$$\begin{aligned}\mathbb{E}[y_i|x_i] &= \mu_i \\ \text{und } \mu_i &= F(\mathbf{x}_i'\boldsymbol{\beta})\end{aligned}$$

zerlegen und  $\eta_i := \mathbf{x}_i'\boldsymbol{\beta}$  setzen, erkennen wir, dass sich der dritte Teil der Definition aus Abschnitt 2.3 als

$$\mu_i = F(\mathbf{x}_i'\boldsymbol{\beta}) \Leftrightarrow \mu_i = H(\eta_i) \Leftrightarrow \eta_i = F^{-1}(\mu_i) = G(\mu_i)$$

ergibt. D.h. die Binary-Choice-Modelle in der obigen Formulierung sind als Spezialfall mit der Annahme einer Bernoulli-Verteilung an  $y$  und einer noch zu wählenden Responsefunktion  $H(x) = F(x)$  in der allgemeinen Definition enthalten.

### 2.4.3. Modelle mit latenten Variablen

Binary Choice-Modelle sind neben der Modellierung von direkt beobachtbaren Größen vor allem bei der Modellierung von **latenten** (d.h. nicht direkt, sondern nur durch Indikatoren beobachtbaren) Variablen von Bedeutung.

#### Mikroökonomische Entscheidung mit latenten Variablen

Greene gibt in [Gre08b] Kap. 23.3.2 das Beispiel einer typischen mikroökonomischen Entscheidung: Ein Konsument erwirbt ein Produkt, wenn der zusätzliche Nutzen durch eine weitere Einheit (der Grenznutzen) seine Kosten für diese weitere Einheit (die Grenzkosten) übersteigt. Dieses ist äquivalent dazu, dass die Differenz von Grenznutzen und Grenzkosten positiv ist. Die konkrete Höhe dieser beiden Variablen bzw. der Differenz ist jedoch nicht beobachtbar. Es ist lediglich beobachtbar, ob der Konsument eine weitere Einheit erworben hat oder nicht.

Allgemein sei ein linearer Zusammenhang zwischen der latenten Variablen  $y^*$  und beobachtbaren Regressoren  $\mathbf{x}$  gegeben:

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \quad (2.7)$$

Die Verteilungsfunktion des Störterms  $\epsilon$  sei durch  $F(x)$  gegeben und es sei der Zusammenhang zwischen der beobachtbaren (Indikator-) Variablen  $y$  und der latenten Variablen  $y^*$  durch

$$y = \begin{cases} 1 & \text{falls } y^* > 0 \\ 0 & \text{falls } y^* \leq 0 \end{cases}$$

gegeben. Im obigen Beispiel bedeutet dieses ein Konsument kauft das Produkt ( $y = 1$ ), wenn die Differenz aus Grenznutzen und Grenzkosten positiv ist ( $y^* > 0$ ), andernfalls erwirbt er es nicht ( $y = 0$ ).

Aus den Überlegungen

$$\begin{aligned} \mathbb{P}(y^* > 0 | \mathbf{x}) &= \mathbb{P}(\mathbf{x}'\boldsymbol{\beta} + \epsilon > 0 | \mathbf{x}) \\ &= \mathbb{P}(\epsilon > -\mathbf{x}'\boldsymbol{\beta} | \mathbf{x}) \end{aligned} \quad (2.8)$$

$$\begin{aligned} &= 1 - \mathbb{P}(\epsilon \leq -\mathbf{x}'\boldsymbol{\beta} | \mathbf{x}) \\ &= 1 - F(-\mathbf{x}'\boldsymbol{\beta}) \end{aligned} \quad (2.9)$$

erkennen wir, dass die Modellspezifikation auf das grundlegende Discrete-Choice-Modell zurückgeführt werden kann. In dieser Modellierung gibt die Verteilungsfunktion der stochastischen Störterme  $F(x)$  dann unmittelbar die Responsefunktion des Modells an.

Unter der Annahme einer symmetrischen Verteilungsfunktion lässt sich (2.9) zu

$$\mathbb{P}(y^* > 0 | \mathbf{x}) = F(\mathbf{x}'\boldsymbol{\beta})$$

vereinfachen.

#### 2.4.4. Random Utility Modelle

Als einen dritten Kontext in dem Discrete Choice-Modelle in der Ökonometrie auftreten nennt [Gre08b] Kap. 23.3.3 die Random Utility Modelle. Im Gegensatz zu den Modellen aus Abschnitt 2.4.3 wird in diesen Modellen der Nutzen zwischen zwei Alternativen  $a$  und  $b$  betrachtet und die Alternative mit dem höheren Nutzen vom Individuum gewählt. Analog zu Abschnitt 2.4.3 ist der Nutzen eines spezifischen Produktes nicht direkt beobachtbar, sondern nur die Entscheidung für oder wider einer der Alternativen.

Wir wollen im Folgenden lineare Nutzenmodelle der Form

$$\begin{aligned} U_a &= \mathbf{x}'\boldsymbol{\beta}_a + \epsilon_a \\ U_b &= \mathbf{x}'\boldsymbol{\beta}_b + \epsilon_b \end{aligned}$$

betrachten.

Für die Indikatorvariable  $y$  gelte

$$y = \begin{cases} 1 & \text{falls } U_a > U_b \\ 0 & \text{falls } U_a \leq U_b. \end{cases}$$

Durch die Überlegungen

$$\begin{aligned} \mathbb{P}[y = 1|\mathbf{x}] &= \mathbb{P}[U_a > U_b] \\ &= \mathbb{P}[\mathbf{x}'\boldsymbol{\beta}_a + \epsilon_a - \mathbf{x}'\boldsymbol{\beta}_b - \epsilon_b > 0|\mathbf{x}] \\ &= \mathbb{P}[\epsilon_a - \epsilon_b > -\mathbf{x}'(\boldsymbol{\beta}_a - \boldsymbol{\beta}_b)] \\ &= 1 - F(-\mathbf{x}'\boldsymbol{\beta}) \end{aligned}$$

mit  $\epsilon := \epsilon_a - \epsilon_b$ ,  $F(x)$  als Verteilungsfunktion von  $\epsilon$  und  $\boldsymbol{\beta} := \boldsymbol{\beta}_a - \boldsymbol{\beta}_b$  ist zu erkennen, dass auch in diesem Modellrahmen die Wahrscheinlichkeit für eine Alternative auf die Verteilungsfunktion der Störterme  $\epsilon$  und damit auf ein Discrete-Choice-Modell zurückgeführt werden kann.

#### 2.4.5. Parameterschätzung

Zur genauen Untersuchung des modellierten Zusammenhangs ist es nötig die Parameter  $\boldsymbol{\beta}$  des Modells zu schätzen. Dieses wird mit der Maximum-Likelihood-Methode (siehe A.1) durchgeführt. Da  $y$  nur die beiden Werte 0 oder 1 annimmt (bzw. Bernoulli-verteilt ist), gilt für die Verteilungsfunktion

$$f(y) = (F(\mathbf{x}'\boldsymbol{\beta}))^y (1 - F(\mathbf{x}'\boldsymbol{\beta}))^{1-y}.$$

Für  $n$  unabhängige Beobachtungen  $\mathbf{x}_i, y_i$  ergibt sich hieraus die Likelihood-Funktion

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n ((F(\mathbf{x}_i'\boldsymbol{\beta}))^{y_i} (1 - F(\mathbf{x}_i'\boldsymbol{\beta}))^{1-y_i})$$

und die log-Likelihood-Funktion

$$\ln(l(\boldsymbol{\beta})) = \sum_{i=1}^n y_i \ln(F(\mathbf{x}_i'\boldsymbol{\beta})) + (1 - y_i) \ln(1 - F(\mathbf{x}_i'\boldsymbol{\beta})). \quad (2.10)$$

Als notwendige Bedingung für die Likelihood-Schätzung erhalten wir hieraus

$$\frac{\partial \ln l}{\partial \beta} = \sum_{i=1}^n y_i \frac{f(\mathbf{x}_i' \beta)}{F(\mathbf{x}_i' \beta)} + (1 - y_i) \frac{-f(\mathbf{x}_i' \beta)}{1 - F(\mathbf{x}_i' \beta)} = 0$$

mit  $f(x)$  als Dichtefunktion von  $F(x)$ .

Anhand der Gleichung (2.10) können wir erkennen, dass zur Maximum-Likelihood-Schätzung die Auswertung der Verteilungsfunktion  $F(x)$  an verschiedenen Stellen erforderlich ist. Falls  $F(x)$  nicht analytisch darstellbar ist, kann die Auswertung von  $F(x)$  an diesen Stellen nur approximativ erfolgen und die zur Maximierung nötige Likelihood-Funktion kann ebenfalls nur approximativ bestimmt werden.

Dieses Prinzip wird auch als Maximum-Simulated-Likelihood-Schätzung (MSL-Schätzung) bezeichnet.

### 2.4.6. Beispiele für Verteilungsfunktionen

In den bisherigen Überlegungen wurde keine spezifische Annahme bezüglich der Verteilungsfunktion, bzw. Responsefunktion  $F(x)$  gemacht. Gemäß [Gre08b] Kap. 23.3.1, [Woo02] Kap. 15.3 und [MN89] Kap. 4.1.4 sind die folgenden Spezifikationen von  $F(x)$  in der ökonometrischen Modellierung von Bedeutung.

#### Probit-Modell

Im Probit-Modell wird angenommen, dass die Störterme  $\epsilon$  (in der latenten Formulierung des Modells) standard-normalverteilt sind, d.h. die Responsefunktion  $F(x)$  ist gegeben durch die Verteilungsfunktion einer standardisierten Normalverteilung

$$F(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

#### Logit-Modell

Im Logit-Modell wird die Responsefunktion spezifiziert als die einfache Logistik-Funktion

$$F(x) = \Lambda(x) = \frac{\exp(x)}{1 + \exp(x)} = \frac{1}{\exp(-x) + 1}.$$

#### Gumbel-Modell

Im Gumbel-Modell wird die Responsefunktion spezifiziert als die Verteilungsfunktion der „Extreme-Value-Type 1“-Verteilung

$$F(x) = \exp(-\exp(-x)).$$

### Komplementäres log-log-Modell

In diesem Modell ist die Responsefunktion durch die Verteilungsfunktion der „Extreme-Minimum-Value“-Verteilung

$$F(x) = 1 - \exp(\exp(-x))$$

gegeben.

Anhand von Abbildung 2.1 erkennt man die Symmetrie der Dichtefunktionen der Probit- und Logit-Modelle.

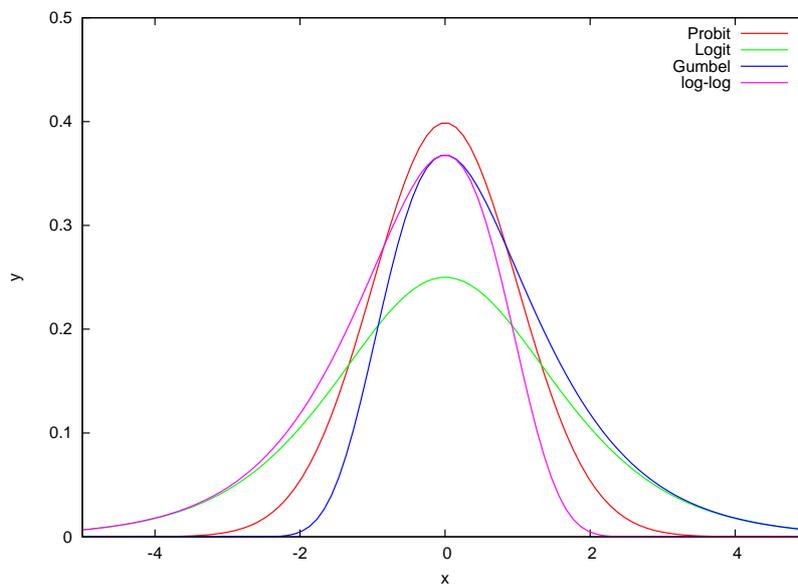


Abbildung 2.1.: Dichtefunktionen von verschiedenen Responsefunktionen

## 2.5. Modelle für Panel-Daten

Bevor wir in den nächsten beiden Abschnitten die multikategorialen Erweiterungen der beiden wichtigsten Binary-Choice-Modelle (Probit und Logit) vorstellen erläutern wir in diesem Abschnitt einige wichtige Definitionen zur Klassifikation von Modellen mit Panel-Daten. Panel-Datensätze bestehen sowohl aus Querschnittsdaten als auch aus Längsschnittdaten, d.h. es werden mehrere Stichproben (Querschnitt) an unterschiedlichen Zeitpunkten (Längsschnitt) von ein oder mehr Variablen untersucht. Z.B. könnte das Haushaltseinkommen von mehreren Familien in verschiedenen Jahren betrachtet werden.

Panel-Daten tauchen in den Erweiterungen der Probit- und Logit-Modelle in den Abschnitten 2.6 und 2.7 wieder auf und haben durch die Zunahme von zur Verfügung stehenden Panel-Datensätzen - und damit einhergehend die Auswertung und Schätzung dieser - in der Ökonomie an Bedeutung gewonnen. Beispiele für derartige Datensätze sind u.a. das „German Socioeconomic Panel“ oder das „Michigan Panel Study of Income Dynamics“, die gesundheitsökonomische bzw. arbeitsmarktökonomische Daten beinhalten.

Diese Art von Modellen lässt sich durch die Gleichungen

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\alpha} + \epsilon_{it}, & i = 1, \dots, n, t = 1, \dots, T_i \\ &= \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \epsilon_{it}, & i = 1, \dots, n, t = 1, \dots, T_i \end{aligned} \quad (2.11)$$

beschreiben. Hierbei ist  $\mathbf{x}_{it} \in \mathbb{R}^{K \times 1}$  ein Vektor, der die Beobachtungen von  $K$  abhängigen Variablen bei  $n$  Individuen zu  $T_i$  Zeitpunkten beinhaltet. Im Speziellen wird eine Konstante nicht zugelassen.

Der Index  $i$  bei  $T_i$  deutet an, dass jedes Individuum  $i$  unterschiedlich häufig beobachtet werden konnte. [Gre08b] Kap. 9.2.4 unterscheidet hier den Fall von **balancierten** Daten ( $T_i = T, i = 1, \dots, n$ ) und **unbalancierten** ( $T_i \neq T_j$ ) Daten.

Der Vektor  $\boldsymbol{\beta} \in \mathbb{R}^{K \times 1}$  erklärt den globalen Einfluss von  $\mathbf{x}_{it}$  auf  $y_{it}$  über alle Individuen und Zeitpunkte hinweg. Demgegenüber steht ein individualspezifischer Effekt  $\mathbf{z}_i'\boldsymbol{\alpha} = c_i$ , der z.B. vom Geschlecht und der ethnischen Abstammung abhängt, über alle Zeitpunkte  $T_i$  konstant ist und je nach Modellzusammenhang beobachtbar oder auch nicht beobachtbar sein kann. [Gre08b] Kap. 9.2.1 spezifiziert diesen Effekt als die **Heterogenität** und führt hierauf aufbauend die folgende Klassifikation von Modellen für Panel-Daten ein:

1. **Gepoolte Regression:** Es gilt  $\mathbf{z}_i = z_i$ , d.h.  $\mathbf{z}_i$  enthält nur einen konstanten Term und das Panel-Daten-Modell kann als ein lineares Regressionsmodell mit einer Konstante  $\alpha$  interpretiert werden.
2. **Fixed-Effects:** Die individuellen  $\mathbf{z}_i$  sind nicht beobachtbar aber mit  $\mathbf{x}_{it}$  korreliert. Das führt dann auf ein Modell der Form (2.11), in dem  $c_i$  als ein unbeobachtbarer, individualspezifischer Term betrachtet wird. Gemäß Greene in [Gre08b] Kap. 9.2.1 kann dieses Modell durch die Formulierung

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \epsilon_{it}$$

beschrieben werden. Dabei wird  $\alpha_i$  als eine individualspezifische Konstante geschätzt. Der Ausdruck „Fixed-Effects“ steht hier lediglich als Synonym dafür, dass eine Korrelation, ein *fixer* Zusammenhang, zwischen  $\mathbf{x}_{it}$  und  $c_i$  angenommen wird. Er schließt insbesondere nicht aus, dass  $c_i$  ein stochastischer Effekt ist.

3. **Random Effects:** Auch hier sind die individuellen  $\mathbf{z}_i$  bzw.  $c_i$  nicht beobachtbar, aber auch nicht mit  $\mathbf{x}_{it}$  korreliert. Dieses Modell kann durch

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha + u_i + \epsilon_{it}$$

mit einer globalen Konstante  $\alpha$  und einer individualspezifischen, zufälligen Komponente  $u_i$  dargestellt werden. Fasst man  $\alpha + u_i$  zusammen, so kann dieses Modell auch als ein Modell mit einer stochastischen Konstante interpretiert werden.

4. **Random Parameter:** Hierbei wird von stochastischen Parametern ausgegangen. Diese Modellierung wird durch

$$y_{it} = \mathbf{x}_{it}'(\boldsymbol{\beta} + \mathbf{h}_i) + (\alpha + u_i) + \epsilon_{it}$$

ausgedrückt. Der Vektor  $\mathbf{h}_i$  gibt die zufällige Störung in den Parametern  $\boldsymbol{\beta}$  für das Individuum  $i$  an.

## 2.6. Probit-Modelle

Während im Abschnitt 2.4 die Darstellung des univariaten Discrete-Choice-Modells und die Einbettung in die Theorie der generalisierten linearen Modelle im Vordergrund steht, wollen wir in diesem Abschnitt verschiedene Erweiterungen des univariaten Probit-Modells aus der Ökonometrie betrachten.

Unser Fokus liegt dabei auf den Erweiterungen für die bei der Maximum-Likelihood-Schätzung der Modellparameter die Auswertung der multivariaten Normalverteilung erforderlich ist. Wir erläutern die Definition jedes Modells und zeigen auf für welche Modellierungen es geeignet ist. Ferner geben wir für jedes Modell die Likelihood-Funktion an und geben eine Referenz für eine konkrete Untersuchung mit diesem Modell an.

### 2.6.1. Panel Probit-Modell

Eine Modell-Spezifikation, die es nötig macht multivariate Verteilungen auszuwerten, sind Probit-Modelle für Panel-Daten. Diese untersuchen die **binäre** Entscheidung von  $n$  Individuen zu  $T_i$  Zeitpunkten und sind daher als eine Kombination der binären Modelle in Abschnitt 2.4 und des allgemeinen Modells für Paneldaten in Abschnitt 2.5 zu sehen. Eine Spezifikation dieser Modelle ist die Annahme, dass ein Parametervektor  $\beta$  für alle Zeitpunkte und Individuen gilt. Bertschek et al. nennen dieses Modell in [BL98] das „Panel Probit Model“. Es lässt sich für unabhängige Stichproben  $i = 1, \dots, n$  durch

$$y_{it}^* = \mathbf{x}_{it}'\beta + \epsilon_{it}, \quad t = 1, \dots, T,$$

$$y_{it} = \begin{cases} 1 & \text{falls } y_{it}^* > 0 \\ 0 & \text{sonst} \end{cases}$$

beschreiben.

Der Vektor  $\mathbf{x}_{it} \in \mathbb{R}^{K \times 1}$  enthält hierbei die Beobachtungen zu Individuum  $i$  zum Zeitpunkt  $t$ . Der Vektor  $\beta \in \mathbb{R}^{K \times 1}$  beschreibt die Partialeffekte, den Einfluss einer Änderung, von  $\mathbf{x}_{it}$  auf die Variablen  $y_{it}^*$ .

Des Weiteren erkennen wir, dass es sich um ein latentes Modell handelt, d.h. dass die  $y_{it}^*$  nicht direkt, sondern nur durch die Indikatorvariablen  $y_{it}$ , beobachtbar sind.

Zur Schätzung der Parameter dieses Modells ist es nötig die Likelihood- bzw. die log-Likelihood-Funktion zu bestimmen. Die Likelihood-Funktion ergibt sich in diesem Modell für ein Individuum  $i$  als

$$L_i(\beta | \mathbf{y}_i, \mathbf{X}_i) = \mathbb{P}[y_{i1}, \dots, y_{iT} | \mathbf{X}_i, \beta] = \int_{L_{iT}}^{U_{iT}} \dots \int_{L_{i1}}^{U_{i1}} f(\epsilon_{i1}, \dots, \epsilon_{iT}) d\epsilon_{i1} \dots d\epsilon_{iT},$$

mit den Entscheidungen von Individuum  $i$  zu allen Zeitpunkten  $\mathbf{y}_i \in \mathbb{R}^{1 \times T}$  und den zugehörigen  $K$  Regressoren zu allen  $T$  Zeitpunkten  $\mathbf{X}_i \in \mathbb{R}^{1 \times T \times K}$ .

Die Grenzen des Integrals sind analog zur univariaten Formulierung des latenten Modells in

Gleichung (2.8) durch

$$U_{it} = \begin{cases} +\infty & \text{falls } y_{it} = 1 \\ -\mathbf{x}_{it}'\boldsymbol{\beta} & \text{falls } y_{it} = 0 \end{cases}$$

$$L_{it} = \begin{cases} -\mathbf{x}_{it}'\boldsymbol{\beta} & \text{falls } y_{it} = 1 \\ -\infty & \text{falls } y_{it} = 0 \end{cases}$$

gegeben.

Durch die Annahme, dass sich jedes Individuum unabhängig von den anderen Individuen entscheidet, können wir die Likelihood-Funktion als

$$L(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}) = \mathbb{P}(\mathbf{y}_1, \dots, \mathbf{y}_n | \mathbf{X}, \boldsymbol{\beta}) = \prod_{i=1}^n \mathbb{P}(\mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\beta})$$

$$= \prod_{i=1}^n L_i(\boldsymbol{\beta} | \mathbf{y}_i, \mathbf{X}_i)$$

und die log-Likelihood-Funktion als

$$\log(L(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X})) = \sum_{i=1}^n \log(L_i(\boldsymbol{\beta} | \mathbf{y}_i, \mathbf{X}_i))$$

ableiten. Beide Funktionen liegen offenbar in der Klasse  $C^\infty(\Omega)$ . Die Dimension des Integrationsproblems ist durch die Anzahl der Zeitpunkte  $T$  gegeben.

Je nach Annahme an die Verteilung der Störterme  $\epsilon_{it}$  ist die Auswertung der Likelihood-Funktion in diesem Modell unterschiedlich schwierig. In jedem Fall sind die  $\epsilon_{it}$  normalverteilt.

Butler et al. spezifizieren in [BM82] eine Erweiterung des Random Effects-Modells, das durch zusätzliche Annahmen an die Störterme  $\epsilon_{it}$  eine geschickte Reduktion des Integrationsproblems auf eine eindimensionale Integration erlaubt. Die Grundannahme des Modells ist eine Zerlegung der Störterme  $\epsilon_{it}$  in einen individualspezifischen Term  $\mu_i$  und einen Term  $\nu_{it}$

$$\epsilon_{it} = \mu_i + \nu_{it}$$

mit

$$\epsilon_{it} \sim \mathcal{N}(0, \sigma^2) \text{ und } \text{Corr}(\epsilon_{it}, \epsilon_{is}) = \rho.$$

Dieses ermöglicht über die Verwendung der Unabhängigkeit der auf den individualspezifischen Term  $\mu_i$  bedingten  $\epsilon_{it}$  die Reduktion auf eine Dimension. Greene zeigt in [Gre08b] Kap. 23.5.1, dass diese Modellformulierung auch eine beliebige, aber konstante Varianz von  $\mu_i$  zulässt.

Da diese Modellformulierung die Modellierungsmöglichkeiten dennoch erheblich einschränkt, werden wir im Folgenden nur das Panel Probit-Modell mit einer allgemeinen Kovarianzstruktur

$$(\epsilon_{i1}, \dots, \epsilon_{iT})' \sim \mathcal{N}_T(\mathbf{0}, \boldsymbol{\Sigma})$$

betrachten. Die Kovarianzmatrix  $\boldsymbol{\Sigma}$  ist hier für jedes Individuum gleich. Dieser allgemeine Ansatz schließt insbesondere auch die Formulierung von Butler et al. ein.

Eine Beispieluntersuchung ist in [Ber95] zu finden. Bertschek analysiert hier den Zusammenhang zwischen Produktinnovationen und Importen sowie ausländischen Investitionen bei deutschen Firmen über einen Zeitraum von 5 Jahren.

### 2.6.2. Multivariate Probit-Modelle

Eine sehr ähnliche Struktur wie das Panel Probit-Modell in Abschnitt 2.6.1 weisen die **multi-variablen** Probit-Modelle auf. Auch sie untersuchen einen Zusammenhang mit latenten Variablen.

Im Gegensatz zu Panel Probit-Modellen werden in multivariaten Probit-Modellen jedoch nicht verschiedene Individuen zu  $T$  verschiedenen Zeitpunkten betrachtet, sondern verschiedene Individuen mit  $M$  **binären** Entscheidungsproblemen konfrontiert. Je nach Anwendung und Verfügbarkeit der Daten können diese Entscheidungen zeitlich parallel oder seriell erfolgen, obwohl die Modellierung die zeitliche Komponente nicht berücksichtigt. In multivariaten Probit-Modellen steht daher die Analyse von Zusammenhängen zwischen den Entscheidungen im Vordergrund.

Ein weiterer, entscheidender Unterschied zum Panel Probit-Modell ist die Annahme, dass für jede Entscheidung spezifische Partialeffekte  $\beta_m$  berücksichtigt werden. Daher kann dieses Modell als ein System von Gleichungen aufgefasst werden in dem jede Gleichung von der Form eines einzelnen univariaten Probit-Modells analog zu Gleichung (2.7) ist. Gemäß Greene in [Gre08b] Kap. 23.9 lässt sich das Modell für die (unabhängigen) Individuen  $i = 1, \dots, n$  als

$$y_{im}^* = \mathbf{x}_{im}'\beta_m + \epsilon_{im} \quad m = 1, \dots, M$$

$$y_{im} = \begin{cases} 1 & \text{falls } y_{im}^* > 0 \\ 0 & \text{sonst} \end{cases}$$

beschreiben.

Dabei gelte für die Störterme  $\epsilon_{im}$  für jedes Individuum  $i$

$$(\epsilon_{i1}, \dots, \epsilon_{iM})' \sim \mathcal{N}_m(\mathbf{0}, \Sigma)$$

mit einer beliebigen, aber für jedes Individuum gleichen, Kovarianzmatrix  $\Sigma$ .

Die Likelihood-Funktion für ein Individuum  $i$  ergibt sich in diesem Modell dann als

$$L_i(\beta_1, \dots, \beta_M | \mathbf{y}_i, \mathbf{X}_i) = \mathbb{P}[\mathbf{y}_i | \mathbf{X}_i, \beta_1, \dots, \beta_M]$$

$$= \int_{L_{iM}}^{U_{iM}} \dots \int_{L_{i1}}^{U_{i1}} f(\epsilon_{i1}, \dots, \epsilon_{iM}) d\epsilon_{i1} \dots d\epsilon_{iM}$$

mit allen Entscheidungen  $\mathbf{y}_i \in \mathbb{R}^{1 \times M}$  für Individuum  $i$  und den zugehörigen Regressoren  $\mathbf{X}_i \in \mathbb{R}^{1 \times M \times K}$ .

Die oberen und unteren Grenzen hängen analog zum Panel Probit-Modell davon ab, welche Werte die Indikatorvariablen  $y_{im}$  annehmen:

$$U_{im} = \begin{cases} +\infty & \text{falls } y_{im} = 1 \\ -\mathbf{x}_{im}'\beta_m & \text{falls } y_{im} = 0 \end{cases}$$

$$L_{im} = \begin{cases} -\mathbf{x}_{im}'\beta_m & \text{falls } y_{im} = 1 \\ -\infty & \text{falls } y_{im} = 0 \end{cases}$$

Daraus ergibt sich dann die globale Maximum-Likelihood-Funktion als

$$L(\beta_1, \dots, \beta_M | Y, \mathbf{X}) = \prod_{i=1}^n L_i(\beta_1, \dots, \beta_M | \mathbf{y}_i, \mathbf{X}_i).$$

Auch hier ist die Likelihood-Funktion wieder beliebig oft differenzierbar und nicht analytisch bestimmbar, d.h. auch in diesem Modell muss einer Maximierung der Likelihood-Funktion, die Simulation derselbigen vorausgehen. Die Anzahl der Wahlmöglichkeiten  $M$  determiniert die Dimension des Integrationsproblems.

### 2.6.3. Multinomiale Probit-Modelle

Eine andere Erweiterung des Probit-Modells ist das multinomiale Probit-Modell, mit welchem das Verhalten von Individuen bei der Wahl zwischen  $J$  verschiedenen Möglichkeiten untersucht werden kann. Es erweitert in dieser Weise die Random Utility Modelle aus Abschnitt 2.4.4 auf Entscheidungsprobleme mit  $J$  verschiedenen Möglichkeiten. In dieser Modellklasse folgen wir der Definition in [Gre08b] Kap. 23.11.5.

Die Modellgleichungen für unabhängige Individuen  $i = 1, \dots, n$  lauten

$$U_{ij} = \mathbf{x}_{ij}'\beta + \epsilon_{ij}, \quad j = 1, \dots, J.$$

Auch in diesem Modell hängt die Schwierigkeit der Parameterschätzung entscheidend von den Annahmen an die Fehlerterme  $\epsilon_{ij}$  ab. Unter der Annahme einer allgemeinen Kovarianzstruktur

$$[\epsilon_{i1}, \dots, \epsilon_{iJ}] \sim \mathcal{N}_J(0, \Sigma),$$

ähnlich wie in den vorhergehenden Abschnitten, ist für die Maximum-Likelihood-Schätzung die Auswertung der durch die Verteilung der Störterme induzierten Verteilungsfunktion einer multivariaten Normalverteilung erforderlich. Dieses ergibt sich aus den folgenden Überlegungen für jedes Individuum  $i = 1, \dots, n$

$$\begin{aligned} \mathbb{P}[y_i = q | \mathbf{X}, \beta] &= \mathbb{P}[U_{iq} > U_{ij}, j = 1, \dots, J, j \neq q] \\ &= \mathbb{P}[\mathbf{x}_{iq}'\beta + \epsilon_{iq} > \mathbf{x}_{ij}'\beta + \epsilon_{ij}, j = 1, \dots, J, j \neq q] \\ &= \mathbb{P}[(\epsilon_{i1} - \epsilon_{iq}) < (\mathbf{x}_{iq}' - \mathbf{x}_{i1}')\beta, \dots, (\epsilon_{iJ} - \epsilon_{iq}) < (\mathbf{x}_{iq}' - \mathbf{x}_{iJ}')\beta] \end{aligned} \quad (2.12)$$

wobei  $y_i = q$  die Wahl von  $q$  durch Individuum  $i$  bedeutet. In Gleichung (2.12) ist die rechte Seite eine multivariate Normalverteilung der Dimension  $J - 1$ , da  $j = q$  nicht berücksichtigt wird.

Wenn wir  $\mathbf{y} = (y_1, \dots, y_n)' \in \mathbb{R}^{n \times 1}$  und  $\mathbf{d} = (d_1, \dots, d_n)' \in \mathbb{R}^{n \times 1}$  definieren, wobei  $d_i$  die Wahl von Individuum  $i$  enthält, ergibt sich die Likelihood-Funktion für die Parameterschätzung als

$$L(\beta | \mathbf{X}, \mathbf{y}) = \prod_{i=1}^n \mathbb{P}[y_i = q_i | \mathbf{X}, \beta].$$

Diese Likelihood-Funktion erfüllt wieder hinreichende Differenzierbarkeitsannahmen und muss zur Maximierung erst approximiert werden.

### 2.6.4. Ordered multivariate Probit-Modelle

Als letzte Erweiterung der multivariaten Probit-Modelle wollen wir eine Erweiterung auf „Ordered Choices“ diskutieren. „Ordered-Choice“-Modelle sind zur Modellierung von Zusammenhängen geeignet, in denen die zu erklärenden Variablen eine implizite Ordnung (z.B. Ratings für Staatsanleihen) aufweisen.

#### Ordered Probit-Modelle

Bevor wir die multivariate Erweiterung der Modelle betrachten, definieren wir die univariaten Ordered Probit-Modelle basierend auf [Gre08b] Kap. 23.10. In diesen Modellen lassen wir  $J + 1$  verschiedene Antwortmöglichkeiten für die beobachtete Indikatorvariable  $y$  zu und definieren diese Modelle werden als latente Modelle durch:

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \quad (2.13)$$

Anders als im univariaten Probit-Modell lassen wir in dieser Modellierung  $J + 1$  verschiedene Antwortmöglichkeiten für die beobachtete Indikatorvariablen  $y$  zu. Insbesondere nehmen wir den Zusammenhang

$$y = \begin{cases} 0 & \text{falls } y^* \leq 0 \\ 1 & \text{falls } 0 < y^* \leq \mu_1 \\ 1 & \text{falls } \mu_1 < y^* \leq \mu_2 \\ \dots & \\ J & \text{falls } \mu_{J-1} \leq y^* \end{cases} \quad (2.14)$$

zwischen den Indikatorvariablen und der latenten Variablen  $y^*$  an. Für die Störterme gilt:

$$\epsilon \sim \mathcal{N}(0, \sigma^2).$$

Die Grenzen der Kategorien  $\boldsymbol{\mu}_{J-1} = (\mu_1, \dots, \mu_{J-1})$  werden dabei als unbekannt vorausgesetzt und müssen zusammen mit den  $\boldsymbol{\beta}$  durch die Maximum-Likelihood-Methode geschätzt werden. Lediglich ihre Anzahl  $J + 1$  ist fest. Die Wahrscheinlichkeitsverteilung für die Indikatorvariable  $y$  bedingt auf die vorhandenen Daten kann dann in einem analogen Vorgehen zu Abschnitt 2.4.3 als

$$\begin{aligned} \mathbb{P}[y = 0 | \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}_{J-1}] &= \Phi(-\mathbf{x}'\boldsymbol{\beta}) \\ \mathbb{P}[y = 1 | \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}_{J-1}] &= \Phi(\mu_1 - \mathbf{x}'\boldsymbol{\beta}) - \Phi(-\mathbf{x}'\boldsymbol{\beta}) \\ \mathbb{P}[y = 2 | \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}_{J-1}] &= \Phi(\mu_2 - \mathbf{x}'\boldsymbol{\beta}) - \Phi(\mu_1 - \mathbf{x}'\boldsymbol{\beta}) \\ &\vdots \\ \mathbb{P}[y = J | \mathbf{x}] &= 1 - \Phi(\mu_{J-1} - \mathbf{x}'\boldsymbol{\beta}) \end{aligned}$$

hergeleitet werden.

Dabei wird vorausgesetzt, dass

$$0 < \mu_1 < \dots < \mu_{J-1}$$

gilt, damit keine der Differenzen negativ ist und so auch tatsächlich eine Wahrscheinlichkeitsverteilung vorliegt.

Auch diese Modelle lassen sich auf den Fall mehrerer Gleichungen des Typs (2.13) bzw. (2.14) erweitern, was dann auf die Definition der multivariaten Ordered Probit-Modelle führt:

Dazu betrachten wir  $n$  Individuen mit  $M$  Entscheidungsproblemen der Form (2.13). Es stehen dabei jeweils  $J_m + 1$  Möglichkeiten für die Indikatorvariablen zur Verfügung. Falls  $y_{im}$  die Indikatorvariable für Individuum  $i$  bezüglich Entscheidung  $m$  ist, ergibt sich das Modell für die Individuen  $i = 1, \dots, n$  und die Entscheidungen  $m = 1, \dots, M$  durch

$$y_{im}^* = \mathbf{x}_{im}'\boldsymbol{\beta}_m + \epsilon_{im} \quad \text{und}$$

$$y_{im} = \begin{cases} 0 & \text{falls } y_{im}^* \leq 0 \\ c & \text{falls } y_{im}^* \in (\mu_{mc-1}, \mu_{mc}], c = 2, \dots, J-1 \\ J & \text{falls } y_{im}^* > \mu_{mJ-1} \end{cases}$$

Lässt man in diesem Modell allgemeine Korrelationen zu, so ergibt sich auch hier für die Bestimmung der relevanten Wahrscheinlichkeiten die Notwendigkeit der Auswertung einer multivariaten Normalverteilung mit allgemeiner Kovarianzmatrix  $\boldsymbol{\Sigma}$ . Das lässt sich durch die Annahme

$$[\epsilon_{i1}, \dots, \epsilon_{iM}] \sim \mathcal{N}_M(0, \boldsymbol{\Sigma}),$$

an die Verteilung der Störterme mit fester Kovarianzmatrix  $\boldsymbol{\Sigma}$  spezifizieren.

Wir verzichten in diesem Modell auf die genaue Angabe der Likelihood-Funktion und geben als ein Beispiel für ein solches Modell die Untersuchung der Zufriedenheit von Touristen in der japanischen Region Hokkaido in [Has10] an. Die implizite Ordnung der Indikatorvariablen ist in diesem Zusammenhang durch Kategorien wie „vollständig zufrieden“, „zufrieden“ etc. gegeben und die multivariate Komponente stellt sich durch mehrere Fragen mit solchen Antwortmöglichkeiten dar.

## 2.7. Logit-Modelle

Neben den in Abschnitt 2.6 angegebenen Probit-Modellen lässt sich auch das Logit-Modell auf ein Modell mit mehreren Wahlmöglichkeiten pro Individuum erweitern. Es erlaubt durch die zum Probit-Modell veränderte Annahme an die Fehlerterme eine alternative Schätzung. Allerdings ist für die Schätzbarkeit des Modells eine Unabhängigkeitsannahme an die Fehlerterme zwingend erforderlich. Diese Annahme, in der Literatur (z.B. [Gre08a] Kap. 0.7.1) auch als Unabhängigkeit von irrelevanten Alternativen (UIA) beschrieben, schränkt die Modellierung von Korrelations- bzw. Abhängigkeitsstrukturen erheblich ein.

Allerdings bilden die **multinomialen** Logit-Modelle den Grundstein für die umfassenden Mixed Logit-Modelle in Abschnitt 2.7.2 und sollen daher kurz dargestellt werden.

### 2.7.1. Multinomiale Logit-Modelle

Die multinomialen Logit-Modelle dienen analog zu den multinomialen Probit-Modellen zur Modellierung von Zusammenhängen in denen Individuen die Wahl zwischen  $J$  verschiedenen Alternativen haben. Dazu gehen wir von einer Modellierung als ein Random Utility-Modell in der

von Greene in [Gre08b] Kap. 23.11.1 vorgeschlagenen Form

$$U_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta}_i + \epsilon_{ij} \quad (2.15)$$

mit Individuen  $i = 1, \dots, n$  aus.

Die Fehlerterme  $\epsilon_{ij}$  folgen unabhängig und identisch-verteilt einer Extremal-1-Verteilung (auch Gumbel-Verteilung) mit Verteilungsfunktion

$$F(\epsilon_{ij}) = \exp(-\exp(-\epsilon_{ij})). \quad (2.16)$$

Des Weiteren sei  $\mathbf{D} \in \mathbb{R}^{n \times J}$  eine Matrix mit  $d_{ij} \in \{0, 1\}$ , die angibt, ob Individuum  $i$  die  $j$  Alternative wählt oder nicht. Da jedes Individuum genau eine Alternative wählt, gilt  $\|\mathbf{D}\|_1 = 1$ .

McFadden zeigt in [McF73], dass sich die Wahrscheinlichkeit für die Wahl von Alternative  $m$  als

$$\mathbb{P}[d_{ij} = 1 | \mathbf{X}_i, \boldsymbol{\beta}_i] = \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\beta}_i)}{\sum_{j=1}^J \exp(\mathbf{x}_{ij}'\boldsymbol{\beta}_i)} \quad (2.17)$$

ergibt.

Darauf aufbauend lässt sich die Likelihood-Funktion mit  $\boldsymbol{\beta} \in \mathbb{R}^{n \times K}$  durch

$$L(\boldsymbol{\beta} | \mathbf{X}, \mathbf{D}) = \sum_{i=1}^n \sum_{j=1}^J \mathbb{P}[d_{ij} = 1 | \mathbf{X}_i, \boldsymbol{\beta}_i]$$

darstellen.

Eine Vereinfachung dieser Modelle sind die **conditionalen** Logit-Modelle, die  $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \dots = \boldsymbol{\beta}_N$  annehmen.

### 2.7.2. Mixed Logit-Modelle

Die im Vergleich zu den Probit-Modellen in Abschnitt 2.6 restriktiven Annahmen an die Verteilung der Störterme haben in der Ökonometrie zur Weiterentwicklung der multinomialen Logit-Modelle geführt. Insbesondere lässt diese Struktur keine Modellierung von Korrelationen für die Störterme  $\epsilon_{ij}$  zu.

Eine Modellform, die dieses z.T. aufhebt, sind Nested Logit-Modelle. Diese heben die Korrelationsbeschränkung durch das Gruppieren von Daten in verschiedene Gruppen auf. Innerhalb der Gruppen wird dann jeweils ein multinomiales Logit-Modell geschätzt. Anschließend werden die Modelle zu einem gesamten multinomialen Logit-Modell kombiniert. Für Details verweisen wir an dieser Stelle auf [Gre08b] Kap. 23.11.4, und stellen stattdessen eine andere Erweiterung, die Mixed Logit-Modelle, in den Vordergrund.

Die Mixed Logit-Modelle sind sehr weitreichendes Framework, die als Random Parameter-Modell gemäß der Einteilung in Abschnitt 2.5 interpretiert werden können.

Durch die Annahme eines stochastischen Störterms ist es möglich die Restriktionen in der Modellierung, die durch die Annahme der Unabhängigkeit von irrelevanten Alternativen (UIA) in

Gleichung (2.16) entstehen, aufzuheben. Dazu wird die Korrelation zwischen den einzelnen Entscheidungen in Gleichung (2.15) über eine Korrelation der  $K$  Regressionsparameter in den  $\beta_i$  realisiert.

Zur formalen Darstellung sei für Individuum  $i$  und die Regressoren  $\beta_i$

$$\beta_i = \theta + v_i$$

gewählt. Hierbei gibt der  $k$ -te Eintrag von  $\theta$ , einen Mittelwert für den  $k$ -ten Regressor für alle  $n$  Individuen an, während  $v_i$  den Störterm für die Verteilung der Parameter angibt.

Die Verteilungsannahmen an die stochastischen Komponenten sind damit die unabhängig und identisch verteilten Störgrößen

$$\epsilon_{ij} \sim \text{Extremal-Type-1-Verteilung}$$

für das eigentliche Modell und die entlang einer Dichtefunktion  $f(\beta_i|\theta)$  verteilten und auf die Mittelwerte  $\theta$  bedingten stochastischen Parameter

$$\beta_i \sim f(\beta_i|\theta).$$

Als Beispiele für in diesem Kontext genutzte Verteilungen für  $\beta_i$  gibt Greene in [Gre08a] Kap. 0.7.3 die (multivariate) Normal- oder die log-Normal-Verteilung an. Die Dichte der zweiten Verteilung mit Mittelwerten  $\mu$  und Kovarianzmatrix  $\Sigma$  ist durch

$$f(\mathbf{x}) = \frac{1}{\sqrt{|\Sigma|(2\pi)^n} \prod_{i=1}^n x_i} \exp\left(-\frac{1}{2} \log(\mathbf{x} - \mu)' \Sigma^{-1} \log(\mathbf{x} - \mu)\right)$$

gegeben, wobei der Logarithmus  $\log(\mathbf{x})$  komponentenweise zu verstehen ist. Weitere mögliche Verteilungen sind laut [Tra03] Kap. 6.1 die Gleichverteilung oder die Dreiecksverteilung.

Aufgrund des Zusammentreffens von einer grundlegenden Logit-Struktur mit einer andersgearteten zusätzlichen stochastischen Verteilungsannahme wird von einem Mixed Modell gesprochen.

Auch Bezeichnungen wie Mixed Multinomial-Modelle sind in der Literatur (z.B. in [HW08]) zu finden und verdeutlichen, dass auch in diesem Kontext eine Modellierung von nicht notwendigerweise binären Entscheidungen möglich ist.

Zur Schätzung des systematischen Teils des Modells  $\theta$  sei  $\mathbf{d}_i \in \mathbb{R}^{J \times 1}$  der Vektor der die Entscheidung für eine der  $j$  Alternativen angibt, d.h. ein Eintrag in  $\mathbf{d}_i$  ist eins und alle anderen sind null. Der Vektor  $\mathbf{x}_i \in \mathbb{R}^{J \times 1}$  enthalte die exogenen Variablen für jede Alternative  $j$  des Individuums  $i$ .

Die Wahrscheinlichkeit, dass die Alternative  $d_i$  gewählt wird, ergibt sich dann als

$$\mathbb{P}^*(\beta_i) = \mathbb{P}[Y_i = d_i | \mathbf{x}_i, \beta_i] = \frac{\prod_{j=1}^J \exp(\mathbf{x}_{ij}' \beta_i)^{d_{ij}}}{\sum_{j=1}^J \exp(\mathbf{x}_{ij}' \beta_i)}. \quad (2.18)$$

Dies ist eine alternative Darstellung der Gleichung (2.17), jedoch enthält diese Wahrscheinlichkeit die zufälligen Parameter  $\beta$ . Dieser stochastische Einfluss kann durch die Überlegungen

$$\mathbb{P}(\theta) = \mathbb{P}[Y_i = d_i | \mathbf{x}_i, \theta] = \int_{\Psi} f(\beta_i|\theta) \mathbb{P}^*[\beta_i] d\beta_i$$

unter Nutzung der Regeln für bedingte Dichtefunktionen herausintegriert werden. Die Integration erfolgt hier über den gesamten Träger  $\Psi$  der Dichtefunktion  $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$  der zugrundeliegenden Zufallsvariablen.

Die Likelihood-Funktion des Problems kann dann durch

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \mathbb{P}(\boldsymbol{\theta})$$

ausgedrückt werden.

Eine Erweiterung dieses Modells ist möglich in dem auch noch eine Zeitkomponente betrachtet wird. Jedes Individuum  $i = 1, \dots, n$  hat die Wahl zwischen  $J$  Alternativen zu  $T$  Zeitpunkten. Da diese Entscheidungen unabhängig voneinander getroffen werden, ergibt sich die Wahrscheinlichkeit aus Gleichung (2.18) hierfür durch

$$\mathbb{P}^*(\boldsymbol{\beta}_i) = \mathbb{P}[Y_i = d_i | \mathbf{x}_i, \boldsymbol{\beta}_i] = \prod_{t=1}^T \frac{\prod_{j=1}^J \exp(\mathbf{x}_{itj}' \boldsymbol{\beta}_i)^{d_{itj}}}{\sum_{j=1}^J \exp(\mathbf{x}_{itj}' \boldsymbol{\beta}_i)},$$

wobei die relevanten Größen  $\mathbf{x}_i$  und  $\boldsymbol{\beta}_i$  auch noch die zeitliche Komponente beinhalten.

Abschließend können wir feststellen, dass auch diese Funktion hinreichende Annahmen an die Differenzierbarkeit erfüllt. Ferner ist in den Mixed Logit-Modellen die Dimension des Integrationsproblems nicht durch die Größe, die durch die multikategoriale Erweiterung, wie die Anzahl der Zeitpunkte  $T$  oder die Anzahl der Alternativen  $J$ , entsteht, sondern durch die Anzahl der Regressoren  $K$  gegeben.



### 3. Hochdimensionale Integration

In diesem Abschnitt stellen wir verschiedene Integrationstechniken zur Approximation eines hochdimensionalen Integranden der Form

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} \approx Qf = \sum_{i=1}^N \omega_i f(\mathbf{x}^{(i)}) \quad (3.1)$$

dar. Dabei sei  $\Omega$  das  $d$ -dimensionale Integrationsgebiet und  $\mathbf{x}^{(i)} \in \Omega$  seien Stützstellen bzw.  $w_i$  Gewichte für die Approximation durch das Quadraturverfahren  $Q$ . Die beschriebenen Quadraturverfahren lassen sich alle als Spezialfälle dieser allgemeinen Darstellung (3.1) auffassen und unterscheiden sich durch die jeweilige Wahl der  $\mathbf{x}^{(i)}$  und  $w_i$ .

Neben der Beschreibung der einzelnen Verfahren wollen wir auch auf die Konvergenzeigenschaften der Verfahren eingehen. Z.B. lässt sich für das Monte Carlo Verfahren eine dimensionsunabhängige Konvergenzrate zeigen, während die Konvergenzrate der anderen Verfahren (Quasi-Monte Carlo, Produktansatz und dünne Gitter) in hohem Maße von der effektiven Dimension des Problems abhängt.

Zur Analyse verwenden wir dabei den absoluten Fehler

$$\epsilon(f) := \left| Qf - \int_{\Omega} f(\mathbf{x}) d\mathbf{x} \right|$$

als ein Maß für die Approximationsgüte (die Konvergenzrate) eines Verfahrens.

#### 3.1. Monte Carlo

Bei der Monte Carlo Integration (MC) werden die Stützstellen  $\mathbf{x}^{(i)}$  als Zufallszahlen einer Gleichverteilung auf dem  $d$ -dimensionalen Integrationsgebiet  $\Omega$  entnommen und mit dem festen Gewicht  $w_i = \frac{1}{N}$  multipliziert. Dieses Verfahren,

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}^{(i)}) \quad \text{mit } \mathbf{x}^{(i)} \sim \mathcal{U}_d(\Omega),$$

führt zu einer dimensionsunabhängigen Konvergenzrate von

$$\epsilon(f) = \mathcal{O}\left(N^{-\frac{1}{2}}\right)$$

unter der Voraussetzung, dass die Varianz beschränkt ist. Dies folgt unmittelbar aus einem Gesetz der großen Zahlen.

Der Nachteil des Verfahrens liegt (gegenüber den anderen Verfahren) in seiner geringen Konvergenzrate für niedrige Dimensionen. Demgegenüber stehen die geringen Voraussetzungen an die Funktion und insbesondere die Dimensionsunabhängigkeit der Konvergenzrate.

Die Verwendung von Monte Carlo Integrationsverfahren, häufig auch Monte Carlo Simulationen genannt, und ihren Erweiterungen wird auch in einer Vielzahl von ökonomischen Artikeln diskutiert. Einen Überblick über verschiedene Methoden gibt Geweke in [Gew96].

#### 3.2. Quasi-Monte Carlo

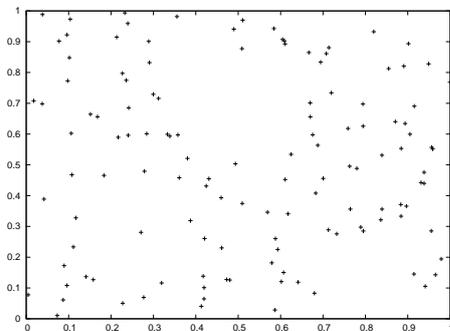
Bei Quasi-Monte Carlo Verfahren (QMC) werden keine Pseudo-Zufallszahlen als Auswertungspunkte des Integranden verwendet, sondern deterministische Punktfolgen sog. Nieder-Diskrepanz-Folgen, die das gegebene Gebiet  $\Omega$  überdecken. Beispiele für solche Folgen auf dem Gebiet  $\Omega = [0, 1]^d$  sind die Sobolev-, Niederreiter- und Halton-Folgen ([Sob76],[Nie78],[Hal60]). Diese Folgen sind „gleichmäßiger“ auf dem Integrationsgebiet verteilt, was sich auch anschaulich an der Gegenüberstellung von MC und QMC für zwei Dimensionen und  $N = 128$  Punkte in Abbildung 3.1 erkennen lässt. Um diese Eigenschaft, die *Diskrepanz*, auch mathematisch exakt zu erfassen sind verschiedene Möglichkeiten entwickelt worden. Für eine Darstellung der Methoden verweisen wir auf den Anhang von [Hol10].

Durch die verbesserten Verteilungseigenschaften kann die Konvergenzrate im geringem und mittleren Dimensionsbereich auf

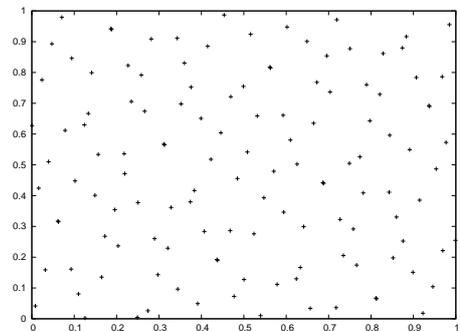
$$\mathcal{O}\left(\frac{\log(N)^d}{N}\right)$$

verbessert werden (siehe z.B. [Nie92]).

Auch in der ökonomischen Literatur finden Quasi-Monte Carlo Verfahren Beachtung. Z.B. diskutiert Bhat in [Bha01] die Verwendung von Sobolev-Folgen für ein Mixed Logit-Modell.



(a) Monte Carlo Gitter



(b) Quasi-Monte Carlo (Sobolev) Gitter

**Abbildung 3.1.:** Integrationsgitter auf dem Gebiet  $[0, 1] \times [0, 1]$  mit  $N = 128$  Punkten

#### 3.3. Produktansatz

Der Produktansatz basiert auf eindimensionalen numerischen Quadraturverfahren. Ein solches Verfahren ist durch

$$Q_l^{(1)} f := \sum_{i=1}^{N_l} w_i f(x_i)$$

mit dem Diskretisierungslevel  $l$  der Quadraturformel gegeben.

Ausgehend von diesem Verfahren wird durch Tensorierung von eindimensionalen numerischen Quadraturformeln eine Quadraturformel für hochdimensionale Integrale konstruiert. Dabei sei  $l_i$  das Diskretisierungslevel der Quadraturformel in der  $i$ -ten Dimension und  $N_{l_i}$  die zugehörige Anzahl von Auswertungspunkten. Dann lässt sich diese Tensorierung schreiben als

$$\begin{aligned} Q_l^{(d)} f &= (Q_{l_1}^{(1)} \otimes \dots \otimes Q_{l_d}^{(1)}) f \\ &= \sum_{i_1=1}^{N_{l_1}} \dots \sum_{i_d=1}^{N_{l_d}} w_{l_1} \cdot \dots \cdot w_{l_d} \cdot f(x_{l_1 i_1}, \dots, x_{l_d i_d}) \end{aligned}$$

mit den  $i_l$ -ten Stützstellen und  $i_l$ -ten Gewichten für eine Quadraturformel mit Level  $l_i$ .

Der Nachteil dieses Ansatzes ist die exponentiell wachsende Anzahl von Auswertungspunkten (der von Bellman in [Bel61] beschriebene „Fluch der Dimension“): Ein Quadraturverfahren mit der gleichen Diskretisierung in jeder Richtung, d.h.  $N := N_{l_1} = \dots = N_{l_d}$ -Auswertungspunkten in jeder Richtung, benötigt in  $d$  Dimensionen insgesamt  $N^d$  Auswertungspunkte.

Die Konvergenzrate des Verfahrens ist gegeben durch

$$\epsilon(f) = \mathcal{O}(n^{-\frac{\alpha}{d}}).$$

Hier gibt  $\alpha$  die Konvergenzrate des eindimensionalen Quadraturverfahrens an.

Da die Konvergenzrate exponentiell von der Dimension  $d$  des Problems abhängt, ist dieses Verfahren für große Dimensionen unbrauchbar.

Wichtige Beispiele für eindimensionale Quadraturverfahren, die auch in den durchgeführten numerischen Experimenten von Bedeutung sind, sollen an dieser Stelle kurz dargestellt werden.

### 3.3.1. Eindimensionale Quadraturverfahren

Aus der Theorie über Quadraturformeln (siehe z.B. [FH07] Kap. 3.6) ist bekannt, dass Gauß-Quadraturformeln den höchsten Exaktheitsgrad ( $2N - 1$  falls  $N$  die Anzahl der Stützstellen beschreibt) erzielen. Bei diesem Verfahren wird die zu integrierende Funktion  $g(x)$  in  $g(x) = \Phi(x)w(x)$  aufgeteilt, wobei der Anteil  $w(x)$  dieser Zerlegung als Gewichtsfunktion bezeichnet wird. Das Integral lässt sich dann durch

$$\int_a^b g(x) dx = \int_a^b \Phi(x)w(x) dx \approx \int_a^b p_n(x)w(x) dx = \sum_{i=1}^n \Phi(x_i)w_i$$

mit einem Polynom  $p_n(x)$  vom Grad  $n$  approximieren.

Damit der optimale Exaktheitsgrad erreicht wird, müssen die Polynome  $p_0(x), \dots, p_n(x)$  eine Orthogonalbasis bzgl. des von der Gewichtsfunktion induzierten Skalarprodukts

$$(f, g)_w := \int_a^b w(x)f(x)g(x)dx$$

bilden. Die Stützstellen  $x_i$  sind dann als Nullstellen des Polynoms  $p_n(x)$  zu wählen und die Gewichte  $w_i$  als Lösung des Gleichungssystems

$$\sum_{i=1}^n p_k(x_i) w_i = \begin{cases} (p_0, p_0) & \text{falls } k = 0 \\ 0 & \text{falls } k = 1, \dots, n-1. \end{cases}$$

Zur Konstruktion der Gewichte und Stützstellen wird in der Praxis häufig ein Zusammenhang zwischen Tridiagonalmatrizen und den Orthogonalpolynomen der jeweiligen Gewichtsfunktion ausgenutzt. Die Stützstellen ergeben sich dann als Eigenwerte dieser Matrizen und die Gewichte als Quadrat des ersten Eintrags der zugehörigen Eigenvektoren. Für eine ausführliche Darstellung verweisen wir auf [FH07] Kap. 3.6 und zitieren lediglich zwei wichtige Beispiele für Gauß-Quadraturformeln aus dieser Quelle.

#### Gauß-Legendre

Die Gauß-Legendre-Quadraturformel ist in ihrer ursprünglichen Form zur Integration auf dem Intervall  $[-1, 1]$  geeignet und hat die Gewichtsfunktion  $w(x) = 1$ . Eine Transformation auf beliebige endliche Intervalle  $[a, b]$  ist durch

$$\int_a^b f(t) dt \approx \frac{b-a}{2} \sum_{i=1}^n w_i f\left(\frac{b-a}{2} x_i + \frac{a+b}{2}\right)$$

mit den ursprünglichen  $x_i$  und  $w_i$  auf  $[-1, 1]$  leicht möglich und erlaubt in dieser Weise die Integration auf beliebigen endlichen Intervallen mit dieser Quadraturformel.

#### Gauß-Hermite

Die zur Gauß-Hermite-Quadraturformel gehörenden Polynome sind orthogonal bezüglich der Gewichtsfunktion  $w(x) = \exp(-x^2)$  und ermöglichen die Integration auf dem gesamten Raum  $\mathbb{R}$ .

#### Clenshaw-Curtis

Neben den Gauß-Quadratur-Formeln ist in dieser Arbeit die Clenshaw-Curtis-Regel aus [CC60] von Bedeutung. Obwohl diese keinen optimalen Exaktheitsgrad hat, ist sie aufgrund der geschachtelten Struktur ihrer Gewichte und Stützstellen (für die später dargestellten dünnen Gitter) sehr hilfreich.

Die Stützstellen und die Gewichte der (offenen) Clenshaw-Curtis-Formeln auf dem Gebiet  $\Omega = (0, 1)$  ergeben sich als

$$x_{li} = \frac{1}{2} \left( 1 - \cos \left( \frac{\pi i}{N_l + 1} \right) \right)$$

und  $w_{li} = \frac{2}{N_l + 1} \sin \left( \frac{\pi i}{N_l + 1} \right) \sum_{j=1}^{(N_l+1)/2} \frac{1}{2j-1} \sin \left( \frac{(2j-1)\pi i}{N_l + 1} \right)$ .

Der Index  $l$  gibt hierbei das Level und  $N_l := 2^l - 1$  die Anzahl der Stützstellen bzw. Gewichte an.

### Generalisierte Gauß-Regel

Eine weitere Möglichkeit ist die Verwendung von generalisierten Gauß-Regeln. Diese versuchen die Integration bei Randsingularitäten durch eine Verschiebung der Singularität in die Gewichtsfunktion zu verbessern. Die Konstruktion einer solchen Regel wird von Oettershagen in [Oet12] vorgestellt.

### 3.4. Dünne Gitter

Um dem „Fluch der Dimension“ bei der regulären Produktintegration zu entgehen, bietet sich der Dünngitter-Ansatz an. Dieses Verfahren hat den Vorteil, dass die Konvergenzrate, ähnlich zu Quasi-Monte Carlo Techniken, nur logarithmisch von der Dimension des Problems abhängt.

Eine Darstellung der dünnen Gitter ist über die Differenz zwischen zwei Quadraturformeln unterschiedlicher Level gegeben. Dazu sei eine Folge von Quadraturformeln mit  $N_l^{(d)}$  Auswertungspunkten für ein Level  $l \in \mathbb{N}$  mit  $N_l^{(d)} < N_{l+1}^{(d)}$  gegeben.

Wir definieren die Differenz zwischen zwei Leveln durch

$$\begin{aligned} \Delta_k^{(1)} f &:= (Q_k^{(1)} - Q_{k-1}^{(1)})f \\ &= \sum_{i=1}^{N_k} w_{k,i} f(x_{k,i}) - \sum_{i=1}^{N_{k-1}} w_{k-1,i} f(x_{k-1,i}) \end{aligned} \quad (3.2)$$

mit  $Q_0 f := 0$ .

Die klassische Dünngitter-Quadratur nach Smolyak (siehe [GG98]) mit  $\mathbf{k} \in \mathbb{N}^d$  ergibt sich dann als

$$Q_l^{(d)} f = \sum_{k_1 + \dots + k_d \leq l + d - 1} (\Delta_{k_1}^{(1)} \otimes \dots \otimes \Delta_{k_d}^{(1)}) f.$$

Es werden also auch hier Tensorprodukte von eindimensionalen Quadraturformeln betrachtet. Anders als im vorhergehenden Produktansatz werden in der Dünngitter-Quadratur nur die Summanden betrachtet, deren Summe der Level echt kleiner als die vorgegebene Konstante  $l + d$  ist. Eine weitere Darstellung, die gemäß Gerstner und Griebel in [GG98] auf Delovs in [Del82] zurückgeht, ist durch die Gleichung

$$Q_l^{(d)} f = \sum_{l \leq \|\mathbf{k}\|_1 \leq l + d - 1} (-1)^{l + d - \|\mathbf{k}\|_1 - 1} \binom{d - 1}{\|\mathbf{k}\|_1 - l} (Q_{k_1}^{(1)} \otimes \dots \otimes Q_{k_d}^{(1)}) f$$

gegeben und wird auch als Kombinationstechnik bezeichnet.

Falls die Quadraturformeln geschachtelt sind, lässt sich die Teleskopsumme in (3.2) zu

$$\sum_{i=1}^{N_k} \tilde{w}_{k,i} f(x_{k,i}) \quad \text{mit} \quad \tilde{w}_{k,i} = \begin{cases} w_{k,i} & \text{falls } i \text{ ungerade} \\ w_{k,i} - w_{k-1, \frac{i}{2}} & \text{falls } i \text{ gerade} \end{cases}$$

### 3. Hochdimensionale Integration

---

zusammenfassen und die Auswertung der eindimensionalen Differenzenformel  $\Delta_k^{(1)}$  benötigt genauso viele Auswertungen wie die Regel  $Q_k$ . Neben der in Abschnitt 3.3.1 dargestellten Clenshaw-Curtis-Regel hat auch die von Gerstner und Griebel in [GG98] analysierte Gauß-Patterson-Regel aus [Pat68] eine solche geschachtelte Struktur. Zu dem erhält diese Regel den erhöhten Exaktheitsgrad der Gauß-Quadraturregeln.

Zur Fehleranalyse des Verfahrens betrachten wir die Klasse  $\mathcal{W}_d^r$  der Funktionen mit beschränkten Ableitungen bis zur Ordnung  $r$ :

$$\mathcal{W}_d^r := \left\{ g : \Omega \rightarrow \mathbb{R}, \left\| \frac{\partial^{|s|} g}{\partial x_1^{s_1} \dots \partial x_d^{s_d}} \right\|_{\infty} < \infty, s_i \leq r \right\}.$$

Erfüllt die eindimensionale Quadraturregel zusätzlich die Fehlerabschätzung

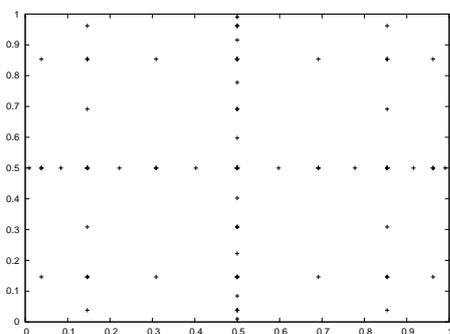
$$\epsilon(f) = \mathcal{O}(n_l^{-r})$$

und gilt für die Auswertungspunkte  $N_l = \mathcal{O}(2^l)$ , dann ergibt sich für alle Funktionen  $f \in \mathcal{W}_d^r$  die Fehlerabschätzung

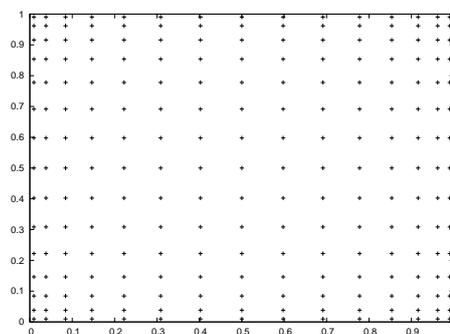
$$\epsilon(f) = \mathcal{O}(N_l^{-r} (\log N_l)^{(d-1)(r+1)})$$

für die Integration mit dünnen Gittern. Dabei wurde eine Gesamtzahl von  $N = \mathcal{O}(N_l (\log N_l)^{d-1})$  Punkten für die Integration verwendet. Der Unterschied der dünnen Gitter für die Clenshaw-Curtis-Regel gegenüber dem Produktansatz ist auch anschaulich in der Abbildung 3.2 zu erkennen.

Die dünnen Gitter können auch in anderen Kontexten als der Integration verwendet werden. Für eine Darstellung solcher Möglichkeiten verweisen wir auf [BG04].



(a) Dünnes Gitter



(b) Volles Gitter

**Abbildung 3.2.:** Integrationsgitter für die Clenshaw-Curtis-Regel für das Level  $l = 4$

## 4. Transformationstechniken

In Abschnitt 2.6 hatten wir gesehen, dass für die Schätzung der Parameter in verschiedenen Arten von Erweiterungen des univariaten Probit-Modells die Berechnung von Wahrscheinlichkeiten der zentrierten multivariaten Normalverteilung,

$$F(\mathbf{a}, \mathbf{b}) = \frac{1}{\sqrt{|\Sigma|(2\pi)^n}} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \exp\left(-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}\right) dx_n \dots dx_1, \quad (4.1)$$

mit beliebiger Kovarianzmatrix  $\Sigma$  erforderlich ist.

Auch wenn eine direkte Anwendung der in Kapitel 3 behandelten Methoden auf dieses Integral in einigen Fällen zu einem Ergebnis führen kann, wird die Konvergenz der behandelten Methoden durch geeignete Transformationen verbessert. Insbesondere erlauben die dargestellten Transformationen auch eine Behandlung von Fällen mit unbeschränkten Gebieten, d.h.  $a_j = -\infty$  oder  $b_k = \infty$  für einige  $j, k \in \{1, \dots, n\}$ , was im Allgemeinen mit den vorher dargestellten Methoden nicht möglich ist.

Alle dargestellten Methoden (mit Ausnahme von Schervishs Methode) wurden ursprünglich für die Verwendung von Monte Carlo Integrationsmethoden entwickelt. Wir werden alle dargestellten Methoden in diesem Abschnitt jedoch vorrangig als Transformationsmethoden auffassen, die dann die Verwendung einer beliebigen Integrationsmethode aus Kapitel 3 erlauben.

Neben einem Ansatz, der eine Transformation auf ein sphärisches Koordinatensystem ermöglicht, stellen wir eine Transformation auf den Einheitswürfel  $(0, 1)^d$  in den Vordergrund und zeigen verschiedene Möglichkeiten zur Verbesserung dieser Transformation auf. Ferner zeigen wir eine in [Hei10] und [RZ07] entwickelte Methode für effizientes hochdimensionales Importance Sampling auf. Für eine umfassende Darstellung der existierenden Methoden verweisen wir auf [GB09].

### 4.1. Cholesky-Zerlegung

Eine zentrale Idee in jedem der im Folgenden dargestellten Verfahren ist die Verwendung der Cholesky-Zerlegung zur Transformation des Integrals (4.1). Dazu wird in einem ersten Schritt die Cholesky-Zerlegung  $\Sigma = \mathbf{L}\mathbf{L}'$  der Kovarianzmatrix  $\Sigma$  berechnet.

Im zweiten Schritt wird eine lineare Transformation  $\mathbf{x} = \mathbf{L}\mathbf{y}$  mit der berechneten Matrix  $\mathbf{L}$  auf das Problem (4.1) angewendet. Durch die Verwendung dieser Transformation mit

$$\mathbf{x}'\Sigma^{-1}\mathbf{x} = \mathbf{y}'\mathbf{L}'\mathbf{L}'^{-1}\mathbf{L}^{-1}\mathbf{L}\mathbf{y} = \mathbf{y}'\mathbf{y}$$

und  $d\mathbf{x} = |\mathbf{L}|d\mathbf{y} = \sqrt{|\Sigma|}d\mathbf{y}$  und unter Verwendung des Transformationssatzes A.1 lässt sich das

Problem als

$$\begin{aligned}
 F(\mathbf{a}, \mathbf{b}) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) d\mathbf{y} \\
 &= \int_{\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}} \prod_{i=1}^n \phi(y_i) d\mathbf{y} \tag{4.2}
 \end{aligned}$$

$$\begin{aligned}
 &= \int_{\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}} \phi(\mathbf{y}) d\mathbf{y} \\
 &= \int_{\mathbb{R}^n} \mathbf{1}[\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}] \phi(\mathbf{y}) d\mathbf{y} \tag{4.3}
 \end{aligned}$$

darstellen.

Eine sehr einfache Möglichkeit zur Berechnung der gesuchten Wahrscheinlichkeiten in (4.1) lässt sich aus (4.3) ableiten. Unter Verwendung der Monte Carlo Integration führt die Gleichung (4.3) unmittelbar auf den in der ökonomischen Literatur (siehe z.B. [HMR96]) bekannten „Crude-Frequency-Simulator“ (CFS).

## 4.2. Deáks Ansatz

Während die Idee dieses Verfahrens, die Transformation des Integrationsproblems auf ein sphärisches Koordinatensystem, von Deák in [Dea80] entwickelt wurde, folgt diese Darstellung Genz in [GB09] Kapitel 4.1.1.

Bei dieser Transformationstechnik wird zunächst die Cholesky-Zerlegung der Kovarianzmatrix berechnet und das Integral analog zum Vorgehen in Abschnitt 4.1 transformiert.

Eine zweite Anwendung des Transformationssatzes stellt  $\mathbf{y}$  in (4.2) durch den Radius  $r$  und den Richtungsvektor  $\mathbf{z}$ , d.h.  $\mathbf{y} = r\mathbf{z}$  mit  $\|\mathbf{y}\|_2 = 1$  und  $\mathbf{y}'\mathbf{y} = r^2$  für  $r \geq 0$  dar. Dieses verändert das Problem (4.2) zu

$$F(\mathbf{a}, \mathbf{b}) = \int_{\|\mathbf{z}\|=1} \frac{2^{1-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_{r_1(\mathbf{z})}^{r_2(\mathbf{z})} r^{n-1} \exp\left(-\frac{r^2}{2}\right) dr d\mathcal{U}(\mathbf{z}) \tag{4.4}$$

mit  $\mathcal{U}(\mathbf{z})$  als Gleichverteilung auf der  $n$ -dimensionalen Einheitssphäre. Die Integrationsgrenzen sind durch

$$\begin{aligned}
 r_1(\mathbf{z}) &= \min\{r : r \geq 0, \mathbf{a} \leq r\mathbf{L}\mathbf{z} \leq \mathbf{b}\} \\
 \text{und } r_2(\mathbf{z}) &= \max\{r : r \geq 0, \mathbf{a} \leq r\mathbf{L}\mathbf{z} \leq \mathbf{b}\}
 \end{aligned}$$

gegeben.

Um diese Transformation praktisch anwenden zu können ist die Auswertung der Gleichverteilung auf der  $n$ -dimensionalen Einheitssphäre erforderlich. Genz gibt in [GB09] Kap. 4.2.1 eine

Transformation aus [FW94] an, die den  $n-1$ -dimensionalen Einheitswürfel  $(0, 1)^{n-1}$  auf die Einheitskugel transformiert. Falls  $\mathbf{w} \in [0, 1]^{n-1}$ , dann sind die  $\mathbf{z}$  mit  $\|\mathbf{z}\| = 1$  gegeben durch

$$z_{n-2i+2}(\mathbf{w}) = \sin(2\pi w_{n-2i+1}) \sqrt{1 - w_{n-2i}^{\frac{2}{n-2i}} \prod_{j=1}^{i-1} w_{n-2j}^{\frac{1}{n-2j}}}$$

$$\text{und } z_{n-2i+1}(\mathbf{w}) = \cos(2\pi w_{n-2i+1}) \sqrt{1 - w_{n-2i}^{\frac{2}{n-2i}} \prod_{j=1}^{i-1} w_{n-2j}^{\frac{1}{n-2j}}}$$

für  $i = 1, \dots, l$  mit  $l = \lfloor \frac{n}{2} \rfloor - 1$  und enden, falls  $n$  gerade ist, mit

$$z_2(\mathbf{w}) = \sin(2\pi w_1) \prod_{j=1}^l w_{n-2j}^{\frac{1}{n-2j}}$$

$$\text{und } z_1(\mathbf{w}) = \cos(2\pi w_1) \prod_{j=1}^l w_{n-2j}^{\frac{1}{n-2j}},$$

oder enden, falls  $n$  ungerade, mit

$$z_3(\mathbf{w}) = (2w_1 - 1) \prod_{j=1}^l w_{n-2j}^{\frac{1}{n-2j}},$$

$$z_2(\mathbf{w}) = 2 \sin(2\pi w_2) \sqrt{w_1(1-w_1)} \prod_{j=1}^l w_{n-2j}^{\frac{1}{n-2j}}$$

$$\text{und } z_1(\mathbf{w}) = 2 \cos(2\pi w_2) \sqrt{w_1(1-w_1)} \prod_{j=1}^l w_{n-2j}^{\frac{1}{n-2j}}.$$

Die Gleichung (4.4) lässt sich dann, unter der Beachtung, dass die Dichte einer  $\chi_n^2$ -Verteilung mit  $n$  Freiheitsgraden durch

$$\chi_n^2(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} \exp(-\frac{x}{2})}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

gegeben ist, als

$$F(\mathbf{a}, \mathbf{b}) = \int_0^1 \int_0^1 \dots \int_0^1 \chi_n(r_2(\mathbf{z}(\mathbf{w}))) - \chi_n(r_1(\mathbf{z}(\mathbf{w}))) d\mathbf{w}$$

darstellen.

### 4.3. Schervishs Methode

Schervish berechnet in [Sch84] das Integral (4.1) über eine Folge von geschichteten Integralen, die dann mit Hilfe von lokal adaptiven Quadraturformeln gelöst werden. Dabei wird eine

Umordnung der Variablen des Integrationsproblems vorgenommen, so dass die beiden größten Integrationsgebiete mit den genauesten Quadraturregeln berechnet werden.

Da sich die Methode von Schervish in hochdimensionalen Fällen ( $d > 5$ ) als ineffizient herausgestellt hat (siehe [Gen93],[GDS02]), verzichten wir auf eine ausführliche Darstellung. Eine wesentliche Idee des Verfahrens, das Umordnen der Integrationsvariablen, wird jedoch in Abschnitt 4.4.1 wieder aufgegriffen.

#### 4.4. Genz-Algorithmus/GHK-Simulator

Ein häufig verwendeter Ansatz ist es das ursprüngliche Integrationsproblem auf den Einheitswürfel  $[0, 1]^n$  im  $\mathbb{R}^n$  zu transformieren. Dieser Ansatz geht sowohl auf Genz (siehe [Gen92]), die Gruppe Geweke und Hajivassiliou (siehe [Gew89] und [BSH93]) sowie auf Keane (siehe [Kea94]) zurück, die dieses Verfahren unabhängig voneinander entwickelt haben. In der statistischen Literatur werden diese Verfahren häufig als „Genz-Algorithmus“ bezeichnet, während in der ökonomischen Literatur die Bezeichnung „GHK-Simulator“ üblich ist.

Dazu wird das eigentliche Integrationsproblem zunächst analog zu Abschnitt 4.1 auf die Form

$$F(\mathbf{a}, \mathbf{b}) = \int_{\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}} \prod_{i=1}^n \phi(y_i) d\mathbf{y} \quad (4.5)$$

transformiert. Aus  $\mathbf{a} \leq \mathbf{L}\mathbf{y} \leq \mathbf{b}$  lassen sich dann, unter Ausnutzung der Dreiecks-Struktur der Matrix  $\mathbf{L}$  mit den Einträgen  $l_{ij}$ , rekursiv (wie beim Vorwärtseinsetzen bei der Lösung eines Gleichungssystems) Integrationsgrenzen

$$a'_i(z_1, \dots, z_{i-1}) = \frac{a_i - \sum_{j=1}^{i-1} l_{ij}y_j}{l_{ii}}$$

$$b'_i(z_1, \dots, z_{i-1}) = \frac{b_i - \sum_{j=1}^{i-1} l_{ij}y_j}{l_{ii}}$$

ableiten. Diese erlauben die Darstellung von (4.5) als

$$F(\mathbf{a}, \mathbf{b}) = \int_{a'_1}^{b'_1} \phi(y_1) \int_{a'_2(y_1)}^{b'_2(y_1)} \phi(y_2) \dots \int_{a'_n(y_1, \dots, y_{n-1})}^{b'_n(y_1, \dots, y_{n-1})} \phi(y_n) d\mathbf{y}.$$

Als nächster Schritt wird eine Transformation der einzelnen Koordination mit der inversen

Normalverteilung  $y_i = \Phi^{-1}(z_i)$  durchgeführt. Das führt auf die Darstellung

$$\begin{aligned} F(\mathbf{a}, \mathbf{b}) &= \int_{\Phi(a'_1)}^{\Phi(b'_1)} \int_{\Phi(a'_2(\Phi^{-1}(z_1)))}^{\Phi(b'_2(\Phi^{-1}(z_1)))} \dots \int_{\Phi(a'_n(\Phi^{-1}(z_1), \dots, \Phi^{-1}(z_{n-1})))}^{\Phi(b'_n(\Phi^{-1}(z_1), \dots, \Phi^{-1}(z_{n-1})))} dz \\ &= \int_{a''_1}^{b''_1} \int_{a''_2(z_1)}^{b''_2(z_1)} \dots \int_{a''_n(z_1, \dots, z_{n-1})}^{b''_n(z_1, \dots, z_{n-1})} dz \end{aligned} \quad (4.6)$$

$$\text{mit } a''_i(z_1, \dots, z_{i-1}) = \Phi \left( \frac{a_i - \sum_{j=1}^{i-1} l_{ij} \Phi^{-1}(z_j)}{l_{ii}} \right)$$

$$\text{und } b''_i(z_1, \dots, z_{i-1}) = \Phi \left( \frac{b_i - \sum_{j=1}^{i-1} l_{ij} \Phi^{-1}(z_j)}{l_{ii}} \right).$$

Auch wenn die zu integrierende Funktion in (4.6) denkbar einfach ist, ist diese Darstellung für die Verwendung der in Kapitel 3 vorgestellten Methoden ungeeignet, da die Integrationsgrenzen von den Integrationsvariablen  $\mathbf{z}$  abhängen.

Eine weitere Anwendung des Transformationssatzes A.1 mit der linearen Transformation

$$z_i = a_i''' + (b_i''' - a_i''')w_i, \quad dz_i = (b_i''' - a_i''')dw_i$$

und den Grenzen

$$\text{mit } a_i'''(w_1, \dots, w_{i-1}) = \Phi \left( \frac{a_i - \sum_{j=1}^{i-1} l_{ij} \Phi^{-1}(a_j''' + (b_j''' - a_j''')w_j)}{l_{ii}} \right)$$

$$\text{und } b_i'''(w_1, \dots, w_{i-1}) = \Phi \left( \frac{b_i - \sum_{j=1}^{i-1} l_{ij} \Phi^{-1}(a_j''' + (b_j''' - a_j''')w_j)}{l_{ii}} \right)$$

transformiert das Problem nun auf den Einheitswürfel  $[0, 1]^n$ :

$$\begin{aligned} F(\mathbf{a}, \mathbf{b}) &= (b_1''' - a_1''') \int_0^1 (b_2'''(w_1) - a_2'''(w_1)) \\ &\quad \dots \int_0^1 b_n'''(w_1, \dots, w_{n-1}) - a_n'''(w_1, \dots, w_{n-1}) \int_0^1 dw. \end{aligned} \quad (4.7)$$

Diese Darstellung erlaubt die direkte Anwendung von mehrdimensionalen Integrationsmethoden. Auch ist an Gleichung (4.7) zu erkennen, dass sich das Integrationsproblem um eine Dimension reduziert hat, da die rechte Seite der Gleichung nicht von  $w_n$  abhängt und diese Variable daher herausintegriert werden kann.

Um auch Integrationsmethoden, die für den Raum  $\mathbb{R}^n$  gegeben sind, verwenden zu können, führen wir eine weitere Transformation  $w_t = \Phi(z_t)$ ,  $dw_t = \phi(z_t)dz_t$  durch, so dass sich das Integrationsproblem als

$$\begin{aligned} F(\mathbf{a}, \mathbf{b}) &= (b_1''' - a_1''') \int_{-\infty}^{\infty} (b_2'''(\Phi(z_1)) - a_2'''(\Phi(z_1))) \\ &\quad \dots \int_{-\infty}^{\infty} b_n'''(\Phi(z_1), \dots, \Phi(z_{n-1})) - a_n'''(\Phi(z_1), \dots, \Phi(z_{n-1})) \int_{-\infty}^{\infty} \phi(\mathbf{z}) dz \end{aligned} \quad (4.8)$$

schreiben lässt.

#### 4.4.1. Variablen-Umordnungstechnik von Gibson

In [GB09] Kap. 4.1.3 gibt Genz verschiedene Techniken zur Verbesserung des Verfahrens in Abschnitt 4.4 an. Diese Verfahren basieren auf einer Umordnung der Integrationsvariablen.

Eines der von Genz beschriebenen Verfahren geht auf Gibson et al. in [GGE94] zurück und sortiert die Integrationsvariablen so, dass die inneren Integrale im Genz-Algorithmus die größten Erwartungswerte und dass die äußeren Integrale (d.h. diejenigen mit kleinem Index  $i$ ) die kleinsten Erwartungswerte haben. Damit wird eine Umordnung entlang der Wichtigkeit der Dimensionen des Integranden erreicht.

Für die Umordnung wird zunächst die Variable mit dem kleinsten Erwartungswert nach der Transformation bestimmt:

$$i = \arg \min_{1 \leq i \leq n} \left\{ \Phi \left( \frac{b_i}{\sqrt{\sigma_{i,i}}} \right) - \Phi \left( \frac{a_i}{\sqrt{\sigma_{i,i}}} \right) \right\}.$$

Durch das Vertauschen der  $i$ -ten und 1-ten Integrationsgrenzen sowie der dazugehörigen Zeilen und Spalten in der Kovarianzmatrix hat das erste (d.h. das äußere) Integral dann den kleinsten Erwartungswert.

Im nächsten Schritt wird die erste Spalte der Cholesky-Zerlegung

$$\begin{aligned} l_{1,1} &= \sqrt{\sigma_{1,1}} \\ l_{i,1} &= \sigma_{i,1}/l_{1,1} \quad \text{für } i = 2, \dots, n \end{aligned}$$

und der Erwartungswert für  $y_1$  durch

$$y_1 = \frac{1}{\Phi(b_1) - \Phi(a_1)} \int_{a_1}^{b_1} s \phi(s) ds$$

bestimmt. Die Einträge von  $\Sigma$  sind dabei mit  $\sigma_{i,j}$  bezeichnet.

Die weiteren Indizes lassen sich daraus rekursiv bestimmen:

Es seien die  $j - 1$ -Integrationsvariablen bereits gewählt und Approximationen für  $y_1, \dots, y_{j-1}$  gefunden. Die  $j$ -te Integrationsvariable wird dann als

$$i = \arg \min_{j \leq i \leq n} \left\{ \Phi \left( \frac{b_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}} \right) - \Phi \left( \frac{a_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}} \right) \right\} \quad (4.9)$$

gewählt.

Als nächstes werden die Integrationsgrenzen, die Zeilen und Spalten von  $\Sigma$  sowie die entsprechenden Zeilen von  $L$  für die Variablen  $i$  und  $j$  vertauscht.

Das Update für den Cholesky-Faktor wird dann durch

$$\begin{aligned} l_{j,j} &= \sqrt{\sigma_{j,j} - \sum_{m=1}^{j-1} l_{j,m}^2} \\ \text{und } l_{i,j} &= \frac{1}{l_{j,j}} \left( \sigma_{i,j} - \sum_{m=1}^{j-1} l_{j,m} l_{i,m} \right) \quad \text{für } i = j + 1, \dots, n \end{aligned}$$

bestimmt.

Schließlich ergibt sich der Erwartungswert für  $y_j$  als

$$y_j = \frac{1}{\Phi(b'_j) - \Phi(a'_j)} \int_{a'_j}^{b'_j} s \phi(s) ds$$

$$\text{mit } a'_j = \frac{a_j - \sum_{i=1}^{j-1} l_{j,i} y_i}{l_{j,j}}$$

$$\text{und } b'_j = \frac{b_j - \sum_{i=1}^{j-1} l_{j,i} y_i}{l_{j,j}}.$$

#### 4.4.2. Variablen-Umordnungstechnik von Genz

In [GB09] Kap. 4.1.3 gibt Genz eine weitere Möglichkeit zum Umordnen der Integrationsvariablen an. Anstatt eine Sortierung zu wählen, die auf dem Erwartungswert der Integrale entlang einzelner Dimensionen basiert, schlägt Genz eine Sortierung anhand der Größe des Integrationsgebietes vor.

Ein kleineres Integrationsgebiet für die Variable  $w_i$  führt aufgrund der geordneten Struktur der Transformation nach Genz zu einer verringerten Varianz von  $w_j$  für  $j > i$ . Basierend auf diesen Überlegungen schlägt Genz vor, beim Umordnen die nächste Variable so zu wählen, dass die Varianz minimiert wird. Dieses bedeutet, dass der Index für die nächste Variable nicht durch (4.9) sondern durch

$$i = \arg \min_{j \leq i \leq n} \left\{ v \left( \frac{b_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}}, \frac{a_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}} \right) \right\},$$

wobei  $v(a, b)$  als Varianz einer standard-normalverteilten Zufallsvariablen  $X$  auf  $[a, b]$ , d.h. als

$$\begin{aligned} v(a, b) &:= \mathbb{V}[X] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{\int_a^b x^2 \exp\left(-\frac{x^2}{2}\right) dx}{\Phi(b) - \Phi(a)} - \left( \frac{\int_a^b x \exp\left(-\frac{x^2}{2}\right) dx}{\Phi(b) - \Phi(a)} \right)^2 \\ &= 1 + \frac{a\phi(a) - b\phi(b)}{\Phi(b) - \Phi(a)} - \left( \frac{\phi(a) - \phi(b)}{\Phi(b) - \Phi(a)} \right)^2 \end{aligned}$$

definiert wird, gewählt werden sollte.

## 4.5. PA-Simulator

Stern stellt in [Ste92] einen teilweise analytischen („partially analytic“=PA) Ansatz zur Lösung des Integrationsproblems für den Fall  $a_1 = \dots = a_n = -\infty$  vor. In diesem Fall entspricht

das Integrationsproblem (4.1) der Auswertung der Verteilungsfunktion einer Zufallsvariablen  $\mathbf{X} \sim \mathcal{N}_n(0, \mathbf{\Sigma})$ .

Zur Bestimmung des Integrals schlägt Stern eine Zerlegung von  $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$  mit  $\mathbf{Y} \sim \mathcal{N}_n(0, \mathbf{\Lambda})$  und  $\mathbf{Z} \sim \mathcal{N}_n(0, \mathbf{\Sigma} - \mathbf{\Lambda})$  mit einer Diagonalmatrix  $\mathbf{\Lambda}$  vor. Zur Varianzreduktion sollte  $\mathbf{\Lambda} = \sigma \mathbf{I}_n$  gewählt werden, wobei  $\sigma$  der kleinste Eigenwert von  $\mathbf{\Sigma}$  ist.

Die grundlegende Idee besteht nun darin das Integral durch die Fehlerzerlegung für  $\mathbf{Y}$  teilweise analytisch zu lösen. Dieses ist aufgrund der Diagonalgestalt von  $\mathbf{\Lambda}$  und der damit einhergehenden Unabhängigkeit der  $Y_i$  möglich. Hierfür wird die auf  $\mathbf{Z}$  bedingte Wahrscheinlichkeit genutzt. Das neue Integrationsproblem ergibt sich dann durch die Überlegungen

$$\begin{aligned}
 \mathbb{P}[\mathbf{X} < \mathbf{b}] &= \mathbb{P}[\mathbf{Y} < \mathbf{b} - \mathbf{Z}] \\
 &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{P}[\mathbf{Y} < \mathbf{b} - \mathbf{z} | \mathbf{z}] f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
 &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n \mathbb{P}[Y_i < b_i - z_i | z_i] f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
 &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n \Phi\left(\frac{b_i - z_i}{\sigma}\right) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}.
 \end{aligned} \tag{4.10}$$

Eine lineare Transformation  $\mathbf{z} = \mathbf{L}_{\mathbf{Z}} \mathbf{e}$ , analog zu Abschnitt 4.1, ermöglicht es nun die Korrelationen aus  $\mathbf{Z}$  in das Produkt in (4.10) zu transformieren. Hierbei ist  $\mathbf{L}_{\mathbf{Z}} \mathbf{L}'_{\mathbf{Z}} = \mathbf{\Sigma} - \mathbf{\Lambda}$ , d.h.  $\mathbf{L}_{\mathbf{Z}}$  ist der Cholesky-Faktor der Kovarianzmatrix von  $\mathbf{Z}$ . Mit  $l_{ij}$  als Elementen von  $\mathbf{L}_{\mathbf{Z}}$  ergibt sich die zu bestimmende Wahrscheinlichkeit als

$$\begin{aligned}
 \mathbb{P}[\mathbf{X} < \mathbf{b}] &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n \Phi\left(\frac{b_i - (\mathbf{L}_{\mathbf{Z}} \mathbf{e})_i}{\sigma}\right) \phi(\mathbf{e}) d\mathbf{e} \\
 &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n \Phi\left(\frac{b_i - \sum_{j=1}^i l_{ij} e_j}{\sigma}\right) \phi(\mathbf{e}) d\mathbf{e}.
 \end{aligned} \tag{4.11}$$

Die Darstellung (4.11) erlaubt nun die Verwendung von Integrationsmethoden, die eine Integration auf dem gesamten Raum  $\mathbb{R}^n$  vorsehen.

#### 4.6. Effizientes hochdimensionales Importance Sampling

Eine weitere Möglichkeit die Integration zu verbessern ist die Nutzung von Importance Sampling Methoden. Während die zuletzt vorgestellten Methoden in den Abschnitten 4.4 und 4.5 bereits als Importance Sampling aufgefasst werden können (siehe [HMR96]), wollen wir in diesem Abschnitt eine moderne Importance Sampling Methode vorstellen, die speziell für hochdimensionale Integranden geeignet ist und die die vorhergehenden Transformationen erweitert.

In [RZ07] beschreiben Richard und Zhang eine Methode für effizientes Importance Sampling in hochdimensionalen Fällen und wenden sie u.a. auf ein Mixed Logit-Modell an. Heiss greift in [Hei10] diese Methode für Integranden des Typs

$$\int_{\mathbb{R}^n} g(\mathbf{z}) \phi(\mathbf{z}) d\mathbf{z} \tag{4.12}$$

auf. Sowohl die Gleichungen (4.3) als auch (4.11) sind sofort durch diese Methode abgedeckt. Auch auf den Genz-Algorithmus/GHK-Simulator ist, unter Verwendung der Gleichung (4.8), die Methode anwendbar.

Die von Heiss spezifizierte Methode lässt sich ausgehend vom allgemeinen Importance Sampling Ansatz

$$\int_{\mathbb{R}^n} \frac{g(\mathbf{z})\phi(\mathbf{z})}{k(\mathbf{z}, \boldsymbol{\theta})} k(\mathbf{z}, \boldsymbol{\theta}) d\mathbf{z}$$

für (4.12) wie folgt beschreiben:

Falls  $k(\mathbf{z}, \boldsymbol{\theta})$  als sog. Gauß-Sampler,

$$k(\mathbf{z}, \boldsymbol{\theta}) := |\mathbf{L}^{-1}| \phi(\mathbf{L}^{-1}(\mathbf{z} - \mathbf{a}))$$

mit einem Parameter-Vektor  $\boldsymbol{\theta}$  und noch zu bestimmenden  $\mathbf{a}$  und  $\mathbf{L}$  gewählt wird, ergibt sich

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{g(\mathbf{z})\phi(\mathbf{z})}{k(\mathbf{z}, \boldsymbol{\theta})} k(\mathbf{z}, \boldsymbol{\theta}) d\mathbf{z} &= \int_{\mathbb{R}^d} \frac{g(\mathbf{z})\phi(\mathbf{z})}{\phi(\mathbf{L}^{-1}(\mathbf{z} - \mathbf{a}))} \phi(\mathbf{L}^{-1}(\mathbf{z} - \mathbf{a})) d\mathbf{z} \\ &= \int_{\mathbb{R}^d} |\mathbf{L}| \frac{g(\mathbf{a} + \mathbf{L}\mathbf{x})\phi(\mathbf{a} + \mathbf{L}\mathbf{x})}{\phi(\mathbf{x})} \phi(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (4.13)$$

$$= \int_{\mathbb{R}^d} h(\mathbf{x}, \boldsymbol{\theta}) \phi(\mathbf{x}) d\mathbf{x} \quad (4.14)$$

$$\text{mit } h(\mathbf{x}, \boldsymbol{\theta}) = |\mathbf{L}| \frac{g(\mathbf{a} + \mathbf{L}\mathbf{x})\phi(\mathbf{a} + \mathbf{L}\mathbf{x})}{\phi(\mathbf{x})}.$$

Die Umformung in Gleichung (4.13) entsteht dabei durch eine einfache Anwendung des Transformationsatzes A.1 mit der affin-linearen Funktion  $h(\mathbf{z}) = \mathbf{a} + \mathbf{L}\mathbf{x}$ .

Die Idee des Importance Samplings liegt darin ein  $\boldsymbol{\theta}^*$  zu finden, so dass

$$h(\mathbf{x}, \boldsymbol{\theta}^*) \approx c \quad (4.15)$$

für eine Konstante  $c$  gilt. Die Lösung des Integrationsproblems ist dann durch die Konstante  $c$  gegeben, da das Problem auf die Integration der Dichtefunktion einer Zufallsvariablen multipliziert mit  $c$  über ihren gesamten Träger vereinfacht werden kann.

Offenbar ist die Eigenschaft aus Gleichung (4.15) gleichbedeutend mit der Wahl von  $\boldsymbol{\theta}$ , so dass

$$g(\mathbf{z})\phi(\mathbf{z}) \approx c \cdot \phi(\mathbf{L}^{-1}(\mathbf{z} - \mathbf{a})) \quad (4.16)$$

gilt.

Der kritische Punkt in dieser Methode ist die Wahl eines geeigneten  $\boldsymbol{\theta}$ . Heiss definiert zur Bestimmung eines solchen  $\boldsymbol{\theta}$  den Fehler zur perfekten Proportionalität in (4.16) durch

$$\begin{aligned} e(\mathbf{z}, \boldsymbol{\theta}) &:= \ln \left( \frac{g(\mathbf{z})\phi(\mathbf{z})}{\phi(\mathbf{L}^{-1}(\mathbf{z} - \mathbf{a}))} \right) + c \\ &= \ln g(\mathbf{z}) - \frac{1}{2} \mathbf{z}' \mathbf{z} + \frac{1}{2} ((\mathbf{z} - \mathbf{a})' (\mathbf{L}\mathbf{L}')^{-1} (\mathbf{z} - \mathbf{a})) + c. \end{aligned} \quad (4.17)$$

Gemäß [Hei10] zeigen Richard und Zhang in [RZ07], dass die minimale Varianz des Fehlers  $e(\mathbf{z}, \boldsymbol{\theta})$  durch die Minimierung des Ausdrucks

$$\tilde{V}(\boldsymbol{\theta}) = \sum_{i=1}^R e^2(\mathbf{z}^r, \boldsymbol{\theta}) g(\mathbf{z}) w^r$$

bestimmt werden kann. Die  $\mathbf{z}^r$  und  $w_r$  geben dabei die Auswertungspunkte bzw. die Gewichte der zur Approximation benutzten Verfahren aus Kapitel 3 an. Eine Minimierung dieses Ausdrucks ist durch eine Erweiterung der kleinsten Quadrate-Schätzung aus Abschnitt 2.2.2 auf lineare Funktionen mit Gewichten, der gewichteten kleinsten Quadrate Regression (siehe [Gre08b] Kap. 8.6), möglich.

Wir definieren  $\mathbf{B}^{-1} := (\mathbf{L}\mathbf{L}')^{-1}$  mit den Einträgen  $b_{jk}$  und schreiben (4.17) zu

$$\ln g(\mathbf{z}) - \frac{1}{2} \mathbf{z}' \mathbf{z} = \alpha + \sum_{j=1}^n \sum_{k=1}^n \beta_{jk} z_j^r z_k^r + \mathbf{z}' \boldsymbol{\gamma} + e(\mathbf{z}^r, \boldsymbol{\theta}) \quad (4.18)$$

um, wobei  $\alpha := c - \frac{1}{2} \mathbf{a}' \mathbf{B}^{-1} \mathbf{a}$ ,  $\beta_{jk} = -\frac{1}{2} b_{jk}$  und  $\boldsymbol{\gamma} := \mathbf{B}^{-1} \mathbf{a}$  definiert wurde.

Die Schätzwerte  $\hat{\alpha}$ ,  $\hat{\beta}_{jk}$  und  $\hat{\boldsymbol{\gamma}}$  für die entsprechenden Größen aus Gleichung (4.18) können gemäß [Hei10] mit Hilfe einer gewichteten linearen Regression geschätzt werden. Wir erhalten dann

$$\begin{aligned} b_{jk} &= -2\hat{\beta}_{jk} \\ \mathbf{a} &= \mathbf{B}\hat{\boldsymbol{\gamma}} \end{aligned}$$

und daraus  $\mathbf{L}$  als Cholesky-Zerlegung von  $\mathbf{B}$  für das Importance Sampling in Gleichung (4.14).

Das Integrationsproblem in der Form von Gleichung (4.14) kann nun mit den Methoden aus Kapitel 3 für den Raum  $\mathbb{R}^n$  approximiert werden.

#### 4.7. Weitere Techniken

Auch wenn in den vorhergehenden Abschnitten einige der wichtigsten Techniken zur Lösung des Problems (4.1) vorgestellt wurden, sind in der Literatur weitere Methoden zur Approximation der gesuchten Wahrscheinlichkeit zu finden. Ein wichtiger Typ unter diesen Methoden sind Ansätze, die das Problem aus der Perspektive der Approximation einer Wahrscheinlichkeit betrachten.

Z.B. beschreiben Gassmann et al. in [GDS02] eine auf Szántai in [Szá76] zurückgehende Approximation, die auf den aus der Stochastik bekannten Bonferroni-Ungleichungen basiert. Da in dieser Arbeit allerdings die Betrachtung des Problems (4.1) als Integrationsproblem im Vordergrund steht, verzichten wir auf eine ausführliche Darstellung dieser Methoden und diskutieren sie nur kurz im Anhang A.3.

## 5. Numerische Resultate

In diesem Kapitel wollen wir die Ergebnisse von Experimenten für die wichtigsten numerischen Integrationsverfahren für die Discrete-Choice-Modelle aufzeigen. Wir gehen dabei sowohl auf die Integrationsprobleme der Probit-Modelle als auch auf diejenigen der Logit-Modelle ein. Für die Konvergenzanalyse der unterschiedlichen Verfahren verwenden wir generierte Daten.

In einem abschließenden Schritt analysieren wir einen Datensatz aus der Gesundheitsökonomie, der erstmals in [RWM03] betrachtet wurde und über die Homepage des *Journal of Applied Econometrics* abgerufen werden kann, mithilfe eines Panel Probit-Modells und der Maximum-Likelihood-Schätzung. Wir verwenden dabei zur Approximation der für die Likelihood-Funktion relevanten Wahrscheinlichkeiten diejenige Kombination aus den analysierten Integrations- und Transformationsmethoden, die sich in den numerischen Experimenten als besonders erfolgreich herausgestellt hat.

### 5.1. Performance-Vergleiche für ein Probit-Modell

Bevor wir die Ergebnisse der Experimente für die Integrale der Probit-Modelle darstellen und erläutern, zeigen wir auf, wie experimentelle Daten modifiziert werden können, so dass die Transformationstechniken für die multivariate Normalverteilung aus Kapitel 4 direkt angewendet werden können. Ferner erläutern wir den zur Analyse verwendeten stochastischen Prozess.

#### 5.1.1. Anpassung der Daten

In Abschnitt 2.6 haben wir gesehen, dass für die Schätzung der Modellparameter in Probit-Modellen Integrale der Form

$$\mathbb{P}[y_1, \dots, y_T | \mathbf{X}] = \int_{L_T}^{U_T} \dots \int_{L_1}^{U_1} f(\epsilon_1, \dots, \epsilon_T) d\epsilon_1 \dots d\epsilon_T$$

mit

$$U_t = \begin{cases} +\infty & \text{falls } y_t = 1 \\ -\mathbf{x}_t' \boldsymbol{\beta} & \text{falls } y_t = 0 \end{cases}$$
$$L_t = \begin{cases} -\mathbf{x}_t' \boldsymbol{\beta} & \text{falls } y_t = 1 \\ -\infty & \text{falls } y_t = 0 \end{cases}$$

zu approximieren sind.

Da wir im Allgemeinen nicht davon ausgehen können, dass  $U_t = \infty$  für alle  $t = 1, \dots, T$  oder  $U_t = -\infty$  für alle  $t = 1, \dots, T$  gilt, ist eine Transformation der Daten nötig, bevor wir die in Kapitel 4

erläuterten Transformationsmethoden für die multivariate Normalverteilung verwenden können. Diese Transformation ist durch

$$\begin{aligned} q_t &:= 2y_t - 1 \\ v_t &:= \frac{1}{\sigma_t} q_t \mathbf{x}_t' \boldsymbol{\beta} \\ (\boldsymbol{\Sigma}^*)_{ij} &:= \frac{q_i q_j}{\sigma_t \sigma_s} (\boldsymbol{\Sigma})_{ij} \end{aligned} \tag{5.1}$$

möglich, wobei  $(\boldsymbol{\Sigma})_{ij}$  die Einträge der ursprünglichen Kovarianzmatrix  $\boldsymbol{\Sigma}$  darstellen.

Nach dieser Transformation gilt

$$\mathbb{P}[y_1, \dots, y_T | \mathbf{X}] = \int_{-\infty}^{v_T} \dots \int_{-\infty}^{v_1} f(\epsilon_1^*, \dots, \epsilon_{e_T}^*) d\epsilon_1^* \dots d\epsilon_T^*,$$

wobei  $f(\epsilon_1^*, \dots, \epsilon_{e_T}^*)$  die Dichtefunktion einer multivariaten Normalverteilung mit der in (5.1) konstruierten Kovarianzmatrix angibt. Die Matrix hat insbesondere normierte Varianzen.

Durch diese Transformationen kann die zu schätzende Wahrscheinlichkeit im Probit-Modell direkt mit den in Kapitel 4 behandelten Methoden approximiert werden.

Für eine Analyse der verschiedenen Integrations- und Transformationsmethoden wollen wir eine in der Ökonometrie (insbesondere in der Zeitreihenanalyse) vielfach benutzte Korrelationsstruktur, den autoregressiven Prozess, verwenden.

### 5.1.2. Autoregressive Korrelationsstruktur

Ein autoregressiver Prozess erster Ordnung, kurz AR(1), ist definiert durch

$$u_t = \rho u_{t-1} + \epsilon_t \text{ für } t \in \mathbb{Z}$$

mit einer Korrelationskonstanten  $\rho \in (-1, 1)$ , sowie unabhängig und identisch verteilten

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Hieraus lässt sich für  $u_t$  die Beziehung

$$u_t = \rho u_{t-1} + \epsilon_t = \rho(\rho u_{t-2} + \epsilon_{t-1}) + \epsilon_t = \dots = \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j}$$

herleiten.

Daraus ergibt sich für den Erwartungswert

$$\mathbb{E}[u_t] = \mathbb{E}\left[\sum_{j=0}^{\infty} \rho^j \epsilon_{t-j}\right] = \sum_{j=0}^{\infty} \rho^j \mathbb{E}[\epsilon_{t-j}] = 0$$

und die Varianz

$$\begin{aligned}
 \mathbb{V}[u_t] &= \mathbb{E}[u_t^2] - \mathbb{E}[u_t]^2 \\
 &= \mathbb{E} \left[ \left( \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j} \right)^2 \right] - 0 \\
 &= \sum_{j=0}^{\infty} \rho^{2j} \mathbb{E}[\epsilon_t^2] = \sum_{j=0}^{\infty} \rho^{2j} \sigma^2 = \frac{\sigma^2}{1 - \rho^2},
 \end{aligned} \tag{5.2}$$

wobei in (5.2) die geometrische Reihe verwandt wird und ausgenutzt wird, dass aus der Unabhängigkeit der  $\epsilon_t$

$$\mathbb{E}[\epsilon_j \epsilon_i] = \mathbb{E}[\epsilon_j] \mathbb{E}[\epsilon_i] = 0 \text{ für } i \neq j$$

folgt.

Das gleiche Argument liefert uns

$$\begin{aligned}
 \mathbb{E}[\epsilon_t u_i] &= \mathbb{E} \left[ \epsilon_t \sum_{j=0}^{i-1} \rho^j \epsilon_{i-j} \right] \\
 &= \sum_{j=0}^{i-1} \rho^j \mathbb{E}[\epsilon_t \epsilon_{i-j}] \\
 &= 0
 \end{aligned}$$

für  $t > i$ .

Dies ermöglicht es den folgenden Ausdruck in (5.3)

$$\begin{aligned}
 cov(u_t, u_{t-\tau}) &= \mathbb{E}[(u_t - \mathbb{E}[u_t])(u_{t-\tau} - \mathbb{E}[u_{t-\tau}])] = \mathbb{E}[(u_t - 0)(u_{t-\tau} - 0)] \\
 &= \mathbb{E} \left[ \left( \sum_{j=0}^{t-1} \rho^j \epsilon_{t-j} \right) u_{t-\tau} \right] = \mathbb{E} \left[ \left( \sum_{j=0}^{\tau-1} \rho^j \epsilon_{t-j} + \sum_{j=\tau}^{t-1} \rho^j \epsilon_{t-j} \right) u_{t-\tau} \right] \\
 &= \mathbb{E} \left[ \left( \sum_{j=0}^{\tau-1} \rho^j \epsilon_{t-j} + \sum_{j=0}^{t-\tau-1} \rho^{j+\tau} \epsilon_{t-\tau-j} \right) u_{t-\tau} \right] = \sum_{j=0}^{\tau-1} \rho^j \mathbb{E}[\epsilon_{t-j} u_{t-\tau}] + \rho^\tau \mathbb{E}[u_{t-\tau}] \\
 &= \rho^\tau \mathbb{V}[u_{t-\tau}] = \rho^\tau \frac{\sigma^2}{1 - \rho^2}
 \end{aligned} \tag{5.3}$$

mit  $\tau \in \{1, \dots, t-1\}$  für die Kovarianz herzuleiten.

Eine wichtige Beobachtung ist, dass sowohl der Erwartungswert als auch die Varianz in jedem Zeitpunkt  $t$  gleich sind, d.h. der Prozess ist stationär.

Es ergibt sich die Kovarianzmatrix für  $T$  aufeinanderfolgende Zufallsvariablen dieses Prozesses

als

$$\Sigma = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}.$$

### 5.1.3. Numerische Ergebnisse

Zum numerischen Vergleich der Verfahren verwenden wir ein latentes Panel Probit-Modell ohne Konstante

$$y_{it} = \mathbf{1}[\beta_1 x_{it} + \epsilon_{it} \geq 0] \text{ mit } t = 1, \dots, T.$$

Die Daten  $x_{it}$  werden dabei durch unabhängig und identisch standard-normalverteilte Zufallsvariablen simuliert. Für die Störterme  $\epsilon_{it}$  verwenden wir einen AR(1)-Prozess

$$\begin{aligned} \epsilon_{it} &= \epsilon_{it-1} + \rho v_{it} \\ \text{mit } v_{it} &\sim \mathcal{N}(0, \sigma^2(1 - \rho^2)). \end{aligned}$$

Dadurch vereinfacht sich die Kovarianzmatrix für die Integration zu

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}.$$

Zum numerischen Vergleich konstruieren wir je 100 Samples für diesen Prozess für verschiedene Kombinationen von  $T$ ,  $\sigma^2$  und  $\rho$ . Die generierten Daten werden dann durch die in Abschnitt 5.1.1 dargestellte Methode auf die zur Integration benötigte Form gebracht. Im letzten Schritt verwenden wir verschiedene Methoden zur Transformation und Auswertung des Integrals.

Weil wir von Daten in „Rohform“ ausgehen und diese vor dem Integrieren anpassen, kann sich auch für gleiche  $T, \sigma^2$  und  $\rho$  das Integrationsproblem leicht verändern, da in der Kovarianzmatrix noch Vertauschungen und Normierungen vorgenommen werden (siehe Abschnitt 5.1.1). Deshalb kann diese Sichtweise auf das Problem als eine ökonometrische angesehen werden, da das ökonometrische Schätzproblem für jedes der Sample gleich ist während sich das numerische Problem leicht verändert.

Da offenbar „korrekte“ Werte nicht verfügbar sind, verwenden wir den durch die Quasi-Monte Carlo Integration mit Sobolev-Folgen und die (einfache) Transformation von Genz ermittelten Wert mit  $n = 10000000$  Auswertungspunkten als „wahren“ Wert für die Konvergenzanalyse der Verfahren.

Als Fehlermaß verwenden wir den gemittelten relativen Fehler

$$\overline{\epsilon_r(f)}_n := \frac{1}{n} \sum_{i=1}^n \frac{|Qf - \int_{\Omega} f(\mathbf{x}) d\mathbf{x}|}{|\int_{\Omega} f(\mathbf{x}) d\mathbf{x}|}.$$

In der Abbildung 5.1 sind die entstehenden Fehler für eine verschiedene Anzahl von Auswertungspunkten bei gleichem  $T$  und verschiedenem  $\sigma^2$  und  $\rho$  mit den unterschiedlichen Methoden dargestellt. Für alle Verfahren ohne den Zusatz „Gibson“ wurde lediglich die Transformation von Genz verwendet und dann eine Approximation des Integrals mit der spezifizierten Methode durchgeführt. Für die Quasi-Monte Carlo (QMC) Experimente wurden Sobolev-Folgen verwendet. Für die Dünngitter-Quadratur wurde die Clenshaw-Curtis-Regel (Spare-CC) bzw. die generalisierte Gauß-Regel (Gauß) verwendet. Bei den Verfahren mit dem Zusatz „Gibson“ verwenden wir zusätzlich die Umordnungstechnik nach Gibson et al.

Die Umordnungstechnik nach Genz wurde nicht mit in die Konvergenzanalyse aufgenommen, da sich in unseren Experimenten gezeigt hat, dass diese nur in sehr wenigen Fällen tatsächlich zu einer signifikanten Umordnung der Variablen führt. Daher ist ein Unterschied in der Approximation gegenüber dem normalen Genz-Algorithmus nur in Ausnahmefällen gegeben, der sich dann nur als sehr gering darstellt.

Die Plots in Abbildung 5.1 zeigen, dass die Korrelationskonstante  $\rho$  maßgeblich die Konvergenz der unterschiedlichen Verfahren beeinflusst. Diese Beobachtung lässt sich durch die Betrachtung der Kovarianzmatrix  $\Sigma$  des Problems erklären. Eine hohe Korrelation führt zu vielen von null verschiedenen Einträgen von  $\Sigma$ , während bei einer geringen Korrelation die Kovarianzmatrix  $\Sigma$  nahezu der Kovarianzmatrix von unabhängigen Zufallsvariablen entspricht, welches das Integrationsproblem zu einem Problem aus Produkten von univariaten Problemen vereinfacht und daher auch eine numerische Approximation deutlich einfacher zu bestimmen ist.

Für niedrige und mittlere Korrelationen führt die Integration mit den dünnen Gittern und der generalisierten Gauß-Regel eindeutig zu den besten Ergebnissen. Erst bei extremer Korrelation verschlechtert sich die Konvergenz der dünnen Gitter soweit, dass das Quasi-Monte Carlo Verfahren den dünnen Gittern überlegen ist.

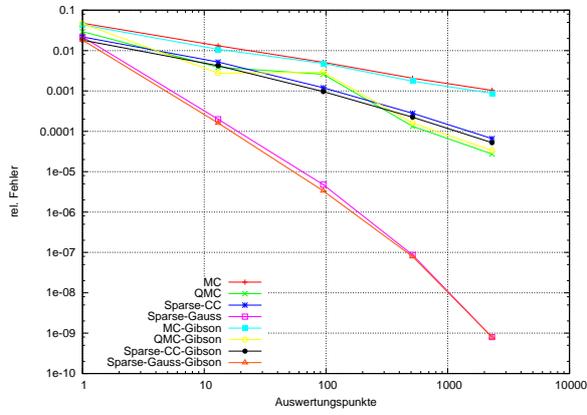
Im Vergleich unterschiedlicher Dimensionen für die gleiche Varianz und die gleiche Korrelation in Abbildung 5.2 wird deutlich, dass der Vorteil der Umordnungstechnik bei hoher Korrelation besonders stark ausgeprägt ist. Dieses lässt sich anhand der Kovarianzmatrix  $\Sigma$  erklären. Für die Kovarianzmatrix eines AR(1)-Prozesses ist eine hohe Korrelationskonstante gleichbedeutend mit einem hohen Quotienten benachbarter Einträge in der Matrix. Dieses wiederum führt zu einem signifikanten Unterschied in der Wichtigkeit der Dimensionen des Integranden. Da die Umordnungstechnik gerade diese Eigenschaft zur Verbesserung der Genz-Transformation ausnutzt, lässt sich so der deutlich stärkere Effekt der Umordnungstechnik für hohe Korrelationen erklären.

Eine weitere Erkenntnis, die sich aus der Abbildung 5.2 gewinnen lässt, ist, dass der positive Einfluss der Umordnungstechnik für große Dimensionen steigt. Sie ist daher besonders zur Verbesserung der Integration in hochdimensionalen Fällen geeignet. Dieses lässt sich dadurch erklären, dass es in hohen Dimensionen besonders wichtig ist die Integrationsdimensionen, die besonders zum Wert des Integrals beitragen, zu identifizieren. Durch die Umordnungstechnik in Kombination mit der geordneten Struktur der Genz-Transformation wird gerade dieser Effekt erreicht.

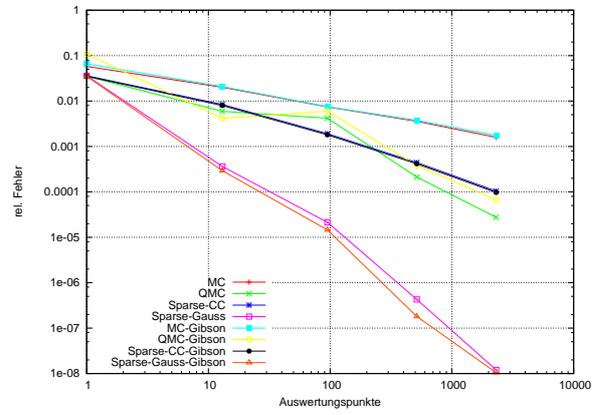
Insgesamt zeigen die numerischen Ergebnisse, dass die Kombination aus Genz-Transformation mit der Umordnungstechnik von Gibson et al. und den dünnen Gittern mit der generalisierten Gauß-Regel auch in hohen Dimensionen zu guten Ergebnissen führt. Lediglich für eine extreme Korrelation sind bei hohen Dimensionen Quasi-Monte Carlo Verfahren zu bevorzugen.

Des Weiteren führt die vorgestellte Umordnungstechnik für kein Integrationsverfahren zu einer Verschlechterung der Konvergenz und sollte daher in Anbetracht des geringen Aufwands ihrer Berechnung im Vergleich zum eigentlichen Integrationsproblem in jeder Integrationsmethode berücksichtigt werden.

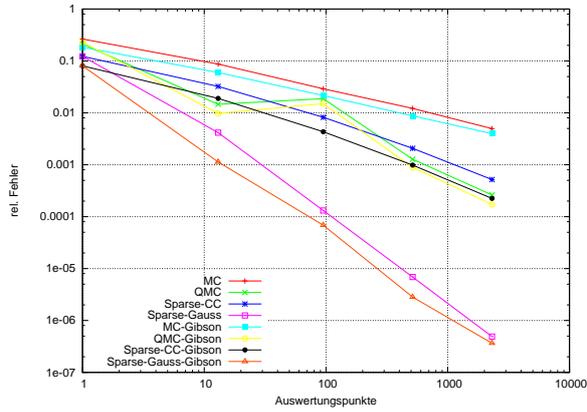
## 5.1. Performance-Vergleiche für ein Probit-Modell



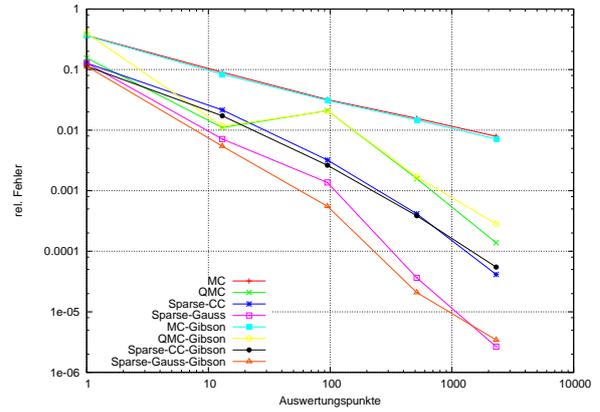
(a)  $\sigma^2 = 1, \rho = 0.1$



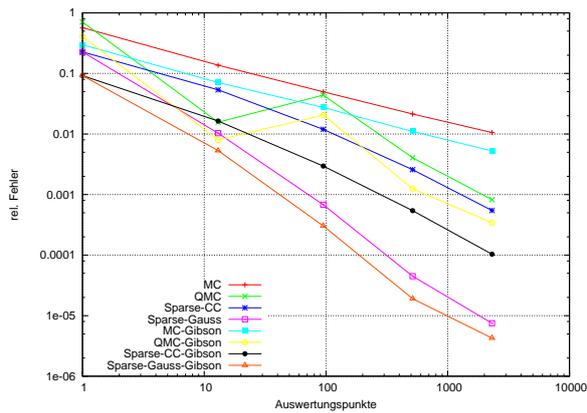
(b)  $\sigma^2 = 5, \rho = 0.1$



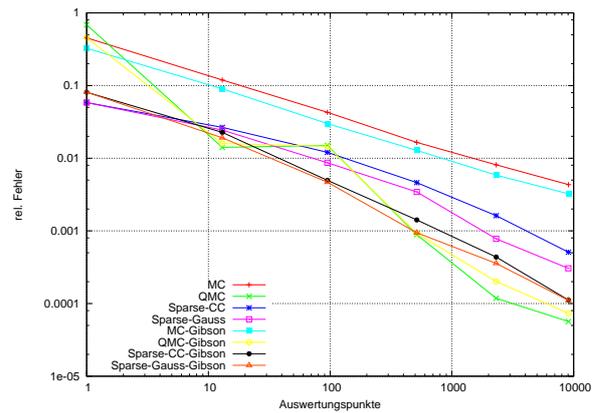
(c)  $\sigma^2 = 1, \rho = 0.5$



(d)  $\sigma^2 = 5, \rho = 0.5$



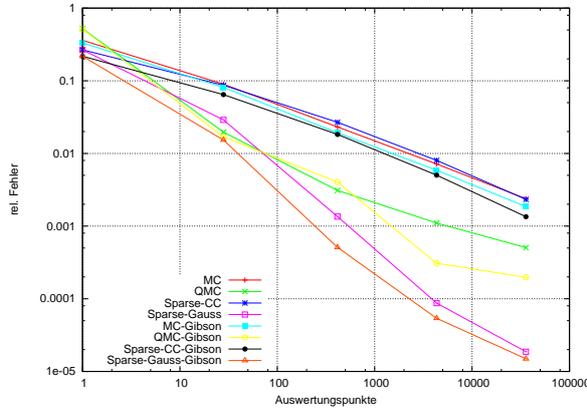
(e)  $\sigma^2 = 1, \rho = 0.9$



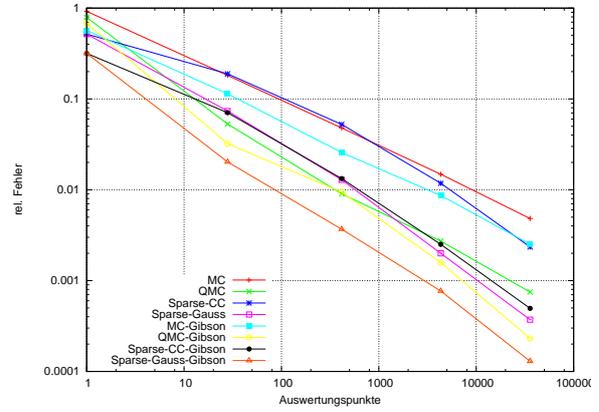
(f)  $\sigma^2 = 5, \rho = 0.9$

Abbildung 5.1.: Der Einfluss von  $\sigma^2$  und  $\rho$  im Probit-Modell mit autoreg. Korrelationsstruktur für  $T = 5$

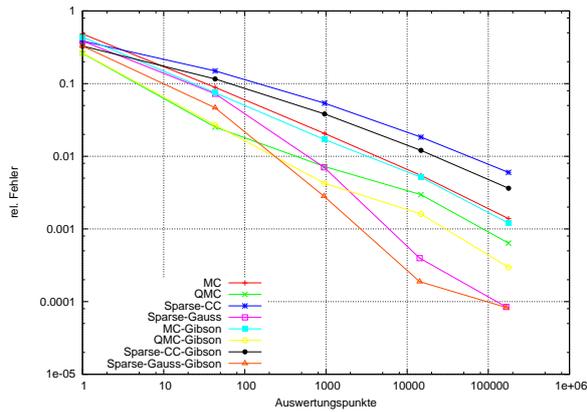
## 5. Numerische Resultate



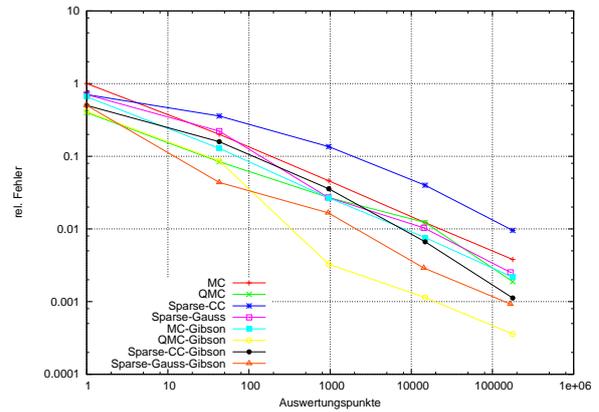
(a)  $T = 10, \sigma^2 = 1, \rho = 0.5$



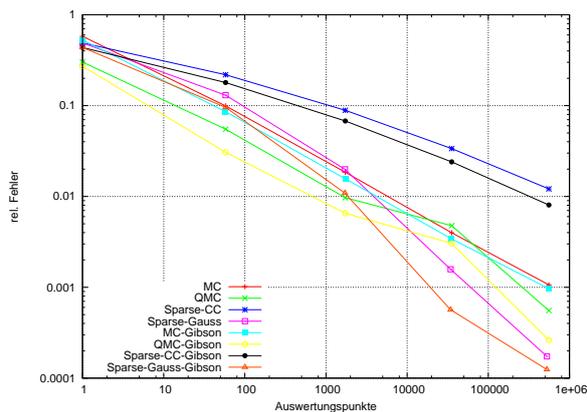
(b)  $T = 10, \sigma^2 = 1, \rho = 0.9$



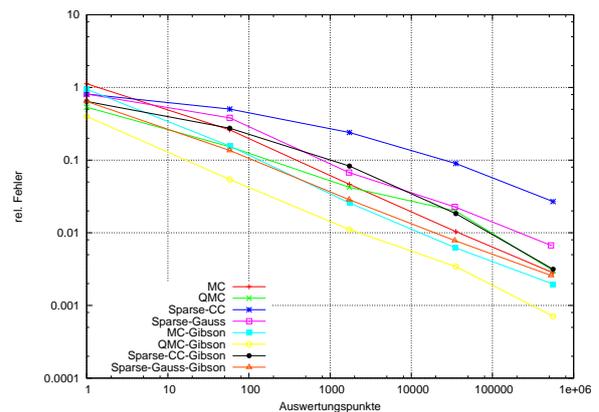
(c)  $T = 15, \sigma^2 = 1, \rho = 0.5$



(d)  $T = 15, \sigma^2 = 1, \rho = 0.9$



(e)  $T = 20, \sigma^2 = 1, \rho = 0.5$



(f)  $T = 20, \sigma^2 = 1, \rho = 0.9$

Abbildung 5.2.: Probit-Modell mit autoreg. Korrelationsstruktur für verschiedene Dimensionen

## 5.2. Anwendung auf ein Mixed Logit-Modell

In diesem Abschnitt wollen wir die in Kapitel 3 vorgestellten Verfahren auf ein Mixed Logit-Modell anwenden. Bei der Analyse steht die Dimension des Integrationsproblems, die in diesen Modellen durch die Anzahl der Regressoren  $K$  gegeben ist, im Vordergrund.

Wie wir in Abschnitt 2.7.2 hergeleitet haben ist das Integrationsproblem in diesem Fall durch

$$\mathbb{P}(\boldsymbol{\theta}) = \mathbb{P}[Y_i = d_i | \mathbf{x}_i, \boldsymbol{\theta}] = \int_{\Psi} f(\boldsymbol{\beta}_i | \boldsymbol{\theta}) \mathbb{P}^*[\boldsymbol{\beta}_i] d\boldsymbol{\beta}_i$$

mit

$$\mathbb{P}^*[\boldsymbol{\beta}_i] = \frac{\prod_{j=1}^J \exp(\mathbf{x}_{ij}' \boldsymbol{\beta}_i)^{d_{ij}}}{\sum_{j=1}^J \exp(\mathbf{x}_{ij}' \boldsymbol{\beta}_i)}$$

gegeben.

Ähnlich zur Untersuchung dieses Modells in [HW08] nehmen wir dabei für die Dichte der Verteilungsfunktion  $f(\boldsymbol{\beta}_i | \boldsymbol{\theta})$  an, dass diese als Produkt von unabhängig und identisch normalverteilten Zufallsvariablen mit Erwartungswert 1 und Varianz  $2/K$  gegeben ist. Des Weiteren simulieren wir unsere Daten entlang der Nutzenfunktion

$$\begin{aligned} U_{ij} &= \mathbf{x}_{ij}' \boldsymbol{\beta}_i + \epsilon_{ij} \\ \text{mit } \epsilon_{ij} &\sim \text{Extremal-Type-1,} \\ \mathbf{x}_{ij} &\sim \mathcal{N}_K \left( 0, \frac{2}{K} \mathbf{1}_K \right), \\ \text{und } \boldsymbol{\beta}_i &= \mathbf{1}_{K \times 1}. \end{aligned}$$

Aus diesen simulierten  $U_{ij}$  wird dann diejenige Alternative mit dem höchsten Nutzen für jedes Sample  $i$  bestimmt und die Entscheidungsmatrix  $\mathbf{D} \in \mathbb{R}^{n \times J}$  mit den Einträgen  $d_{ij}$  entsprechend besetzt.

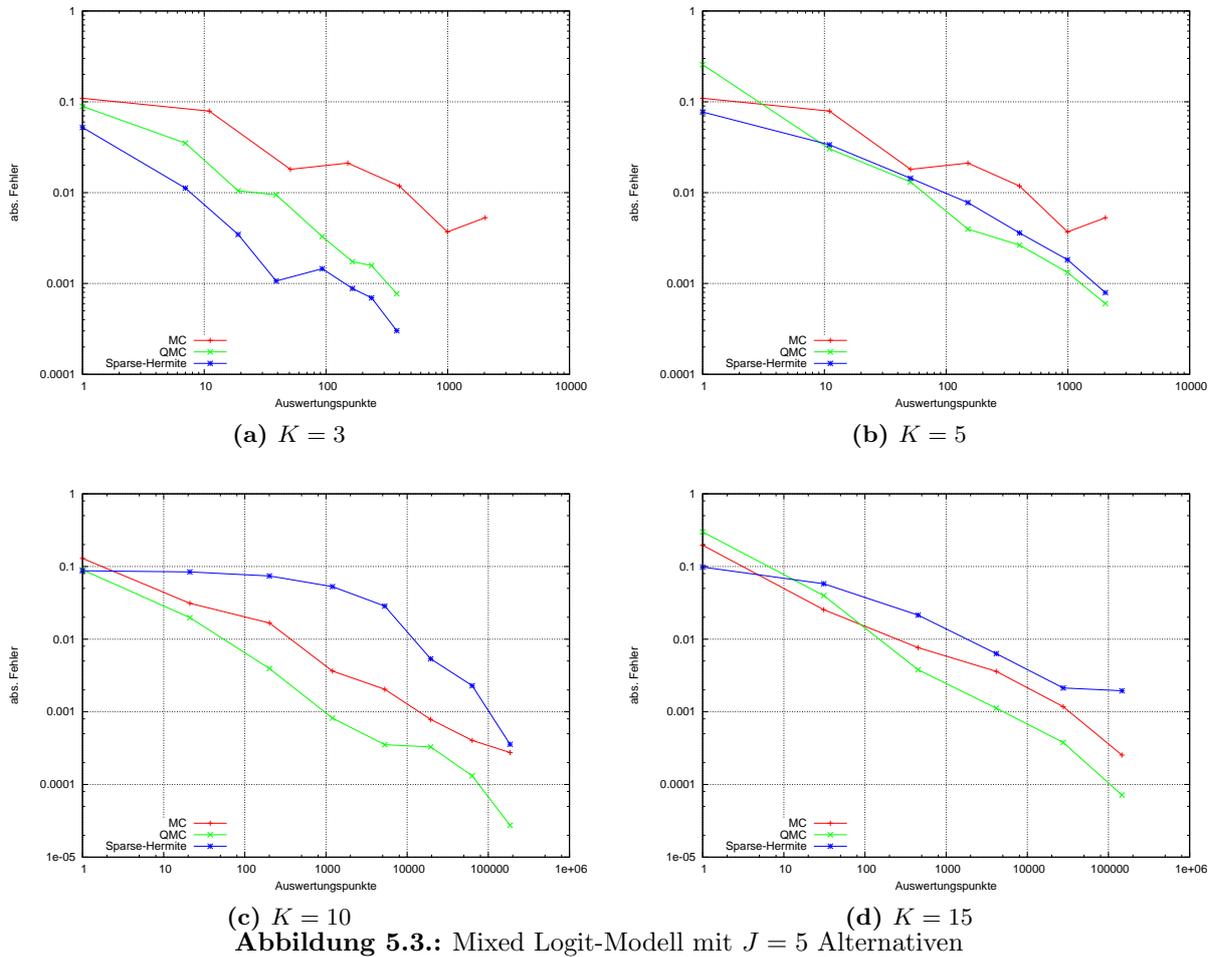
Zur Analyse der Verfahren verwenden wir in diesem Fall den gemittelten *absoluten* Fehler

$$\overline{\epsilon_{abs}(f)}_n := \frac{1}{n} \sum_{i=1}^n \left| Qf - \int_{\Omega} f(\mathbf{x}) d\mathbf{x} \right|$$

und untersuchen die Verfahren für  $n = 100$  Sample.

Neben dem Monte Carlo Verfahren und dem Quasi-Monte Carlo Verfahren (mit Sobolev-Folgen) untersuchen wir dieses Problem für die Dünngitter-Quadratur mit der Gauß-Hermite-Regel (*Sparse* in der Abbildung 5.3) auf dem Raum  $\mathbb{R}$ . Da auch hier „korrekte“ Werte nicht verfügbar sind, verwenden wir wieder eine Approximation mit einer hinreichend großen Anzahl 20000000 von Quasi-Monte Carlo Punkten zum Vergleich.

In Abbildung 5.3 erkennen wir, dass die dünnen Gitter im niedrigdimensionalen Fall  $K = 3$  auch in diesem Modellrahmen die beste Approximation liefern. Jedoch bereits für die Dimension  $K = 5$  führt das Quasi-Monte Carlo Verfahren auf eine ähnlich gute Approximation wie die dünnen Gitter. Für höhere Dimensionen liefert sowohl Quasi-Monte Carlo als auch Monte Carlo bessere Ergebnisse als die Dünngitter-Quadratur.



### 5.3. Panel Probit-Analyse

In diesem Abschnitt analysieren wir die Wahrscheinlichkeit eines Arztbesuchs in Abhängigkeit von den Regressoren Alter, Einkommen, Anzahl der Kinder, Anzahl an Schuljahren und Familienstand mit Hilfe eines Panel Probit-Modells. Das Modell lässt sich durch

$$\begin{aligned}
 \mathbb{P}[\text{ArztBesuch}_{it} > 0] &= F(\beta_1 + \beta_2 \text{Alter}_{it} + \beta_3 \text{Einkommen}_{it} + \beta_4 \text{Kinder}_{it} + \\
 &\quad \beta_5 \text{Schuljahre}_{it} + \beta_6 \text{Familienstand}_{it}) \\
 &= F(\boldsymbol{\beta}' \mathbf{x}_{it})
 \end{aligned}
 \tag{5.4}$$

darstellen. In den Vektoren  $\boldsymbol{\beta}, \mathbf{x}_{it} \in \mathbb{R}^{6 \times 1}$  sind die entsprechenden Größen aus Gleichung (5.4) zusammengefasst. Während die Ausprägungen der Variablen Alter, Einkommen, Kinder und Schuljahre offenkundig sind, ist dieses für den Familienstand nicht der Fall. Dieser Regressor kann die Werte 1 für „verheiratet“ und 0 für „unverheiratet“ annehmen.

Da wir eine Probit-Analyse durchführen ist  $F$  durch die Verteilungsfunktion einer multivariaten Normalverteilung gegeben. Wir spezifizieren die Kovarianzmatrix von  $F$  durch einen AR(1)-Prozess. Dieses scheint plausibel, da z.B. das Alter der gleichen Person in zwei verschiedenen

Jahren offenkundig voneinander abhängt. Da wir die Varianz  $\sigma^2$  und die Korrelation  $\rho$  des Prozesses nicht kennen, werden sie in der Maximum-Likelihood-Schätzung der Parameter integriert.

Als Datensatz, der die genannten Variablen enthält, verwenden wir Ergebnisse von 27.326 Beobachtungen in 7.293 deutschen Haushalten in den Jahren 1984 bis 1995 der über <http://people.stern.nyu.edu/wgreene/Econometrics/healthcare.txt> bzw. über die Homepage des *Journal of Applied Econometrics* abgerufen werden kann. Wir haben aus diesem Datensatz allerdings nur die Beobachtungen verwendet, für die Daten für die gleiche Person zu 7 verschiedenen Zeitpunkten zur Verfügung standen. Zur Schätzung der Parameter  $\beta$  muss die Likelihood-Funktion

$$\log(L(\beta|\mathbf{Y}, \mathbf{X})) = \sum_{i=1}^n \log(L_i(\beta|\mathbf{y}_i, \mathbf{X}_i)) = \sum_{i=1}^n \log(\mathbb{P}[y_{i1}, \dots, y_{iT}|\mathbf{X}, \beta]) \quad (5.5)$$

maximiert werden. Wir verwenden hierzu die Matlab-Routine „fminunc“, wobei die Berechnung von  $\mathbb{P}[y_{i1}, \dots, y_{iT}|\mathbf{X}, \beta]$  mit dem Genz-Algorithmus einschließlich der Umordnung nach Gibson et al. und den dünnen Gittern mit der generalisierten Gauß-Regel erfolgt. Für die benötigten Startwerte dieser Routine machen wir uns zu Nutze, dass Greene in [Gre08b] Bsp. 23.8 das gleiche Problem durch eine wesentlich einfachere Probit-Modellierung analog zu [BM82] untersucht und verwenden die von ihm errechneten Werte für  $\beta, \sigma^2$  als Startwerte für die Matlab-Routine. Falls solche Startwerte nicht zur Verfügung stehen, bietet es sich an sie durch Schätzung eines entsprechenden Modells zu bestimmen. Dieses ist leicht möglich, da in der Likelihood-Funktion in solchen Modellen nur die Auswertung eindimensionaler Verteilungsfunktion erforderlich ist.

Als ein wichtiger Hinweis sei angemerkt, dass wegen der Unabhängigkeit der einzelnen Individuen und damit der einzelnen Summanden in der log-Likelihood-Funktion in 5.5, die Auswertung der Funktion durch Parallelisierung beschleunigt werden kann.

Wir erhalten als Schätzungen für die entsprechenden Parameter im Modell:

$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$
0.1070	1.0797	0.0304	-0.1979	0.2492	0.0482



## 6. Zusammenfassung und Ausblick

### Resultate dieser Arbeit

In dieser Arbeit leiteten wir ausgehend von den grundlegenden Modellen der Ökonometrie, den linearen Regressionsmodellen, die generalisierten linearen Modelle der Statistik her. Wir sahen, dass diese Modelle die Analyse einer Vielzahl von empirischen Datentypen erlauben und konzentrierten uns auf einen besonders wichtigen Spezialfall, die Discrete Choice Modelle. Diese erlauben die Analyse von Entscheidungen zwischen verschiedenen Alternativen in unterschiedlichsten Kontexten.

Wir erläuterten mehrere Erweiterungen der beiden wichtigsten Discrete Choice Modelle (Probit und Logit), die die Modellierungsmöglichkeiten vergrößern und stellten fest, dass in diesen Erweiterungen die Bestimmung von Wahrscheinlichkeiten, die als ein hochdimensionales Integral interpretiert werden können, nötig ist, wenn die Parameter des Modells mit der Maximum-Likelihood-Methode geschätzt werden. Wir erkannten insbesondere, dass in allen multikategorialen Erweiterungen des Probit-Modells die Auswertung der Verteilungsfunktion einer multivariaten Normalverteilung erforderlich ist.

Da die hieraus und die aus den Erweiterungen des Logit-Modells entstehenden Integrationsprobleme nicht analytisch gelöst werden können, sind hierfür approximative Verfahren nötig. Solche Verfahren, beispielsweise Monte Carlo, Quasi-Monte Carlo und dünne Gitter, stellten wir daher im nächsten Kapitel vor. Wir gingen dabei auch auf unterschiedliche eindimensionale Quadraturformeln, die für die dünnen Gitter verwendet werden können, ein.

Wir erkannten, dass zur erfolgreichen Anwendung dieser Methoden auf das Integrationsproblem (die Auswertung der Verteilungsfunktion der multivariaten Normalverteilung) eine Transformation des Problems erforderlich ist und erläuterten daher verschiedene Möglichkeiten einer solchen Transformation. Im Speziellen stellten wir eine u.a. auf Genz in [Gen92] zurückgehende Transformation auf den offenen Einheitswürfel  $(0, 1)^{d-1}$  und eine auf Gibson et al. in [GGE94] zurückgehende Verbesserung dieser Methode vor. Sie ordnet die Integrationsvariablen in der Genz-Transformation entlang ihres Erwartungswerts.

In den numerischen Experimenten in Kapitel 5 konnten wir dann zeigen, dass die Kombination der Genz-Transformation mit der Umordnungstechnik von Gibson et al. zu einer Verbesserung sowohl der Monte Carlo und Quasi-Monte Carlo Integration als auch der dünnen Gitter führt. Wir untersuchten dabei in dieser Arbeit zum ersten Mal die Kombination von dünnen Gittern mit der Umordnungstechnik von Gibson. Wir erkannten, dass der positive Effekt der Umordnungstechnik auf die Genauigkeit der dünnen Gitter in einigen Fällen beträchtlich (bis zu einer Größenordnung) sein kann und beobachteten, dass die Verbesserung dann besonders groß ist, wenn die dünnen Gitter relativ zu den anderen Verfahren besonders schlecht konvergieren. Insgesamt zeigten die numerischen Experimente, dass die Kombination der Genz-Transformation mit der Umordnungstechnik nach Gibson et al. sowie den dünnen Gittern mit generalisierten Gauß-Quadraturregeln in fast allen Fällen am besten konvergiert.

Für die Experimente verwendeten wir einen autoregressiven Prozess erster Ordnung, welcher auch in der praktischen Anwendung in Discrete Choice Modellen verwendet wird, so dass unsere Ergebnisse auch praktische Relevanz aufweisen. Dieses konnten wir an einer Maximum-Likelihood-Maximierung mit realen Daten aus dem Bereich der Gesundheitsökonomie belegen.

### Ausblick

Diese Arbeit bietet Anknüpfungsmöglichkeiten in vielerlei Hinsicht.

Die eine Seite der Anknüpfungspunkte ist die der potenziellen Anwendungen. Hochdimensionale Integrale und insbesondere die Auswertung der Verteilungsfunktion der multivariaten Normalverteilung sind in verschiedenen Zusammenhängen erforderlich, z.B. bei korrelierten Assets im Rahmen des Black-Scholes-Modells in der numerischen Finanzmathematik (siehe z.B. [Gla04]). Weitere Anwendungsmöglichkeiten sind im Bereich des „Stochastic Programming“ (siehe [Pre03]) und in der Bayesschen Statistik (siehe [Sha88]) zu finden.

Auch die Betrachtung des ähnlich gearteten Integrationsproblems der Auswertung der Verteilungsfunktion der zentrierten multivariaten  $t$ -Verteilung

$$T_n(\mathbf{a}, \mathbf{b}, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{|\Sigma|} (\nu\pi)^n} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \left(1 + \frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+n}{2}} d\mathbf{x} \quad (6.1)$$

könnte an die dargestellten Methoden anknüpfen (siehe [GB09]).

Als zweite Seite der Anknüpfungspunkte ist die Verbesserung der eigentlichen numerischen Integrationsmethode zu sehen. Da die von Gibson et al. eingeführte Umordnungstechnik zu einer Umordnung der Integrationsdimension entlang ihres Erwartungswerts führt, führt sie auch eine Ordnung entlang der Wichtigkeit der Dimensionen ein. Hier könnte es sinnvoll sein zu untersuchen, ob diese Priorisierung für die Verwendung von Methoden, die eine solche abfallende Bedeutung von unterschiedlichen Dimensionen ausnutzen können, hinreichend ist. Ein Beispiel sind die von Gerstner und Griebel in [GG03] erläuterten adaptiven dünnen Gitter. Diese können die relevanten Dimensionen des Integrationsproblems erkennen und so die Konvergenz des Verfahrens verbessern.

## A. Anhang

### A.1. Maximum-Likelihood-Schätzung

In diesem Abschnitt stellen wir kurz die Methode der Maximum-Likelihood-Schätzung dar. Wir beschränken uns dabei auf den Spezialfall von „Produktmodellen“, d.h. Modelle, die ausschließlich unabhängige Stichproben der Zufallsvariablen für die Schätzung verwenden. Die Einschränkung ist hinreichend, da alle betrachteten Anwendungen in dieser Arbeit diese Eigenschaft aufweisen. Für eine allgemeine Darstellung verweisen wir auf [Geo09] Kap. 7.3.

Es sei  $f(\mathbf{y}|\boldsymbol{\theta})$  die Dichtefunktion einer  $d$ -dimensionalen Zufallsvariablen bedingt auf die Parameter  $\boldsymbol{\theta}$ . Es seien  $\mathbf{y}_i, i = 1, \dots, n$  unabhängige Beobachtungen dieser Zufallsvariablen, die in der Matrix  $\mathbf{Y} \in \mathbb{R}^{d \times n}$  zusammengefasst sind. Die Likelihood-Funktion  $L(\boldsymbol{\theta}|\mathbf{Y})$  für ein solches Produktmodell ist dann definiert als die Funktion, die den Parametern  $\boldsymbol{\theta}$  die gemeinsame Dichtefunktion zuordnet:

$$L(\boldsymbol{\theta}|\mathbf{Y}) := f(\mathbf{y}_1, \dots, \mathbf{y}_n|\boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{y}_i|\boldsymbol{\theta}).$$

Der Maximum-Likelihood-Schätzer für  $\boldsymbol{\theta}$  ist definiert als das Maximum der Likelihood-Funktion und muss der notwendigen Bedingung

$$\frac{\partial L(\boldsymbol{\theta}|\mathbf{Y})}{\partial \boldsymbol{\theta}} = 0$$

genügen.

Aufgrund der ausschließlich unabhängigen Stichproben  $\mathbf{y}_i$  und der daraus resultierenden multiplikativen Verknüpfung in der Likelihood-Funktion ist es für die Maximierung i.d.R. von Vorteil mit der Log-Likelihood-Funktion, d.h. der logarithmierten Likelihood-Funktion

$$\ln(L(\boldsymbol{\theta}|\mathbf{Y})) = \ln\left(\prod_{i=1}^n f(\mathbf{y}_i|\boldsymbol{\theta})\right) = \sum_{i=1}^n \ln(f(\mathbf{y}_i|\boldsymbol{\theta})),$$

zu arbeiten, da diese dann eine additive Struktur aufweist. Wegen der Monotonie des Logarithmus führt dieses zur selben Schätzung für die Parameter  $\boldsymbol{\theta}$ .

### A.2. Transformationssatz

**Definition A.1.** *Diffeomorphismus:*

*Eine bijektive, stetig differenzierbare Abbildung  $T : U \rightarrow V$  einer offenen Menge  $U \subset X$  auf eine offene Menge  $V \subset Y$  heißt Diffeomorphismus, wenn die Umkehrung  $T^{-1} : V \rightarrow U$  ebenfalls stetig differenzierbar ist.*

**Satz A.1.** *Transformationssatz*

Für jeden Diffeomorphismus  $T : \hat{\Omega} \rightarrow \Omega$ ,  $\hat{\Omega} \subset \mathbb{R}^d$  und jede auf  $\hat{\Omega}$  integrierbare Funktion gilt

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} = \int_{\hat{\Omega}} f \circ T(\mathbf{y}) |\det(DT(\mathbf{y}))| d\mathbf{y}.$$

*Beweis:* Der Beweis ist in vielen Standardwerken über Maß- und Integrationstheorie zu finden, z.B. in [Kön04].

**A.3. Szántais Ansatz**

Ein Ansatz, der sich deutlich von den Integrationsmethoden für Wahrscheinlichkeiten einer multivariaten Normalverteilung unterscheidet schlägt gemäß [GDS02] Szántai in [Szá76] vor. Dieser Ansatz approximiert, im Gegensatz zu den Methoden in Kapitel 4, das eigentliche Problem, die Bestimmung einer Wahrscheinlichkeit und nicht die Bestimmung eines Integrals. Des Weiteren lässt sich dieser Ansatz für die Bestimmung von Wahrscheinlichkeiten auf beliebige multivariate Verteilungen anwenden. Dazu sei das rechteckige Integrationsgebiet  $(\mathbf{a}, \mathbf{b})$  als Schnitt von Halbräumen der Form  $\{\mathbf{x} \in \mathbb{R}^n : x_i \geq a_i\}$  bzw.  $\{\mathbf{x} \in \mathbb{R}^n : x_i \leq b_i\}$  dargestellt. Wenn wir beide Restriktionen zusammenfassen, ergibt sich für das Integrationsgebiet die Darstellung:

$$A := (\mathbf{a}, \mathbf{b}) = \bigcap_{i=1}^n A_i, \text{ mit } A_i = \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq b_i\}$$

Wir definieren, wie üblich,  $\bar{A}_i = \mathbb{R}^n \setminus A_i$  und erhalten dann mit dem Inklusions- und Exklusionsprinzip bzw. der Siebformel die Darstellung

$$\mathbb{P}[A] = 1 - \sum_{i=1}^n \mathbb{P}(\bar{A}_i) + \sum_{i < j} \mathbb{P}(\bar{A}_i \cap \bar{A}_j) - \sum_{i < j < k} \mathbb{P}(\bar{A}_i \cap \bar{A}_j \cap \bar{A}_k) + \dots + (-1)^n \mathbb{P}\left(\bigcap_{i=1}^n \bar{A}_i\right)$$

für die gesuchte Wahrscheinlichkeit  $\mathbb{P}[A]$ . Schneidet man diese Darstellung nach den ersten  $k$  Summanden ab, so erhält man eine untere bzw. obere Abschätzung für die gesuchte Wahrscheinlichkeit. Diese auch als Bonferroni-Ungleichungen bekannten Abschätzungen liefern für  $k = 1$  bzw.  $k = 2$  die Abschätzungen

$$B_1 := 1 - \sum_{i=1}^n \mathbb{P}(\bar{A}_i) \leq \mathbb{P}[A] \leq 1 - \sum_{i=1}^n \mathbb{P}(\bar{A}_i) + \sum_{i < j} \mathbb{P}(\bar{A}_i \cap \bar{A}_j) =: B_2.$$

Da zur Bestimmung von  $\mathbb{P}[\bar{A}_i]$  und  $\mathbb{P}[\bar{A}_i \cap \bar{A}_j]$  lediglich ein- bzw. zweidimensionale Integrale bestimmt werden müssen, lassen sich  $B_1$  und  $B_2$  sehr gut mit numerischen Quadraturverfahren approximieren. Hierauf aufbauend schlägt Szántai die Verwendung der folgenden drei Schätzer für die gesuchte Wahrscheinlichkeit  $\mathbb{P}[A]$  vor:

1.  $\hat{p}_1$  als ein direktes Monte Carlo Sampling des Integrationsproblems  $p$ .
2.  $\hat{p}_2$  als ein Monte Carlo Sampling für das Integral  $p - p_L$ , wobei  $p_L$  die untere Schranke, also  $B_1$ , für die gesuchte Wahrscheinlichkeit angibt. Dieses Sampling wird danach um  $p_L$  korrigiert.

3.  $\hat{p}_3$  als ein Monte Carlo Sampling für das Integral  $p - p_U$ , wobei  $p_U$  die obere Schranke, also  $B_2$ , für die gesuchte Wahrscheinlichkeit angibt. Dieses Sampling wird dann um  $p_U$  korrigiert.

Die Schätzer 2. und 3. lassen sich dann auch als  $\hat{p}_2 = \widehat{p - p_L} + p_L$  und  $\hat{p}_3 = \widehat{p - p_U} + p_U$  darstellen. Zur Varianzreduktion wird schließlich der Ausdruck

$$\hat{p}^{final} = w_1 \hat{p}_1 + w_2 \hat{p}_2 + w_3 \hat{p}_3 \text{ mit } w_1 + w_2 + w_3 = 1$$

minimiert, was den Schätzer  $\hat{p}^{final}$  für  $\mathbb{P}[A]$  liefert.

Eine weniger aufwendige Methode ist die Approximation von  $\mathbb{P}[A]$  durch den Mittelwert  $(\frac{p_U + p_L}{2})$  der errechneten Grenzen  $p_U$  und  $p_L$ . Diese Methode empfiehlt sich aber nur, wenn der absolute Fehler  $\frac{p_U - p_L}{2}$  klein ist.



## Literaturverzeichnis

- [Bel61] BELLMAN, R.: Adaptive Control Processes: A Guided Tour. 1961.
- [Ber95] BERTSCHEK, IRENE: Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investment. *The Journal of Industrial Economics*, 43(4):pp. 341–357, 1995.
- [BG04] BUNGARTZ, HANS-JOACHIM und MICHAEL GRIEBEL: Sparse grids. *Acta Numerica*, 13:1–123, 2004.
- [Bha01] BHAT, C.R.: Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7):677–693, 2001.
- [BL98] BERTSCHEK, IRENE und MICHAEL LECHNER: Convenient estimators for the panel probit model. *Journal of Econometrics*, 87(2):329 – 371, 1998.
- [BM82] BUTLER, J. S. und ROBERT MOFFITT: A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model. *Econometrica*, 50(3):pp. 761–764, 1982.
- [BSH93] BORSCH-SUPAN, A. und V.A. HAJIVASSILOU: Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of econometrics*, 58(3):347–368, 1993.
- [CC60] CLENSHAW, C. W. und A. R. CURTIS: A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205, 1960. 10.1007/BF01386223.
- [Dea80] DEAK, I.: Three digit accurate multiple normal probabilities. *Numerische Mathematik*, 35(4):369–380, 1980.
- [Del82] DELVOS, F.J.: d-Variate Boolean interpolation. *Journal of Approximation Theory*, 34(2):99–114, 1982.
- [FH07] FREUND, R.W. und R.H.W. HOPPE: Stoer/Bulirsch: Numerische Mathematik 1. Springer-Verlag Berlin Heidelberg, 2007.
- [Fri33] FRISCH, R.: Editorial. *Econometrica*, 1:1–4, 1933.
- [FT01] FAHRMEIR, L. und G. TUTZ: Multivariate statistical modelling based on generalized linear models. Springer Verlag, 2001.
- [FW94] FANG, K. und Y. WANG: Number-theoretic methods in statistics, Band 51. Chapman & Hall/CRC, 1994.
- [GB09] GENZ, A. und F. BRETZ: Computation of multivariate normal and t probabilities. Springer Publishing Company, Incorporated, 2009.
- [GDS02] GASSMANN, HI, I. DEÁK und T. SZÁNTAI: Computing multivariate normal probabilities: A new look. *Journal of Computational and Graphical Statistics*, 11(4):920–949, 2002.

- [Gen92] GENZ, A.: Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1:141–149, 1992.
- [Gen93] GENZ, A.: Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, 25:400–400, 1993.
- [Geo09] GEORGII, H.O.: Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik. *de Gruyter*, 2009.
- [Gew89] GEWEKE, J.: Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*, 57:1317–1339, 1989.
- [Gew96] GEWEKE, JOHN: Monte carlo simulation and numerical integration. 1:731–800, 1996.
- [GG98] GERSTNER, T. und M. GRIEBEL: Numerical integration using sparse grids. *Numerical algorithms*, 18(3):209–232, 1998.
- [GG03] GERSTNER, T. und M. GRIEBEL: Dimension-Adaptive Tensor-Product Quadrature. *Computing*, 71(1):65–87, 2003.
- [GGE94] GIBSON, GJ, CA GLASBEY und DA ELSTON: Monte Carlo Evaluation of Multivariate Normal Integrals and Sensitivity to Variate Ordering. *Advances in Numerical Methods and Applications*, 1994.
- [Gla04] GLASSERMAN, P.: Monte Carlo methods in financial engineering, *Band 53*. Springer verlag, 2004.
- [Gre08a] GREENE, W.H.: Discrete choice modeling. *NYU Working Paper No. 2451/26036*, 2008.
- [Gre08b] GREENE, WILLIAM H.: Econometric Analysis. *Pearson Education, Inc.*, 2008.
- [Gri06] GRIEBEL, M.: Sparse grids and related approximation schemes for higher dimensional problems. *Seiten 106–161*, 2006.
- [Hal60] HALTON, J.H.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960.
- [Has10] HASEGAWA, HIKARU: Analyzing tourists' satisfaction: A multivariate ordered probit approach. *Tourism Management*, 31(1):86 – 97, 2010.
- [Hei10] HEISS, F.: The panel probit model: Adaptive integration on sparse grids. *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, 26:41–64, 2010.
- [HMR96] HAJIVASSILIOU, V., D. MCFADDEN und P. RUUD: Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics*, 72(1-2):85–134, 1996.
- [Hol10] HOLTZ, M.: Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance, *Band 77*. Springer Verlag, 2010.
- [HW08] HEISS, FLORIAN und VIKTOR WINSCHER: Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1):62 – 80, 2008.
- [Kea94] KEANE, M.P.: A computationally practical simulation estimator for panel data. *Econometrica: Journal of the Econometric Society*, 62:95–116, 1994.
- [Kön04] KÖNIGSBERGER, K.: Analysis 2. *Springer Verlag*, 2004.

- 
- [McF73] MCFADDEN, D.: Conditional logit analysis of qualitative choice behavior. 1973.
- [MN89] McCULLAGH, P. und J.A. NELDER: Generalized linear models. *Chapman & Hall/CRC*, 1989.
- [Nie78] NIEDERREITER, H.: Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc*, 84(6):957–1041, 1978.
- [Nie92] NIEDERREITER, H.: Quasi-Monte Carlo Methods. *Wiley Online Library*, 1992.
- [Oet12] OETTERSCHAGEN, JENS: Dimension - Adaptive Sparse Grid Quadrature for Integrals with Boundary Singularities. *In: 2nd Workshop on Sparse Grids and Applications Munich*, 2012.
- [Oku62] OKUN, ARTHUR: Potential GNP, its measurement and significance. *Cowles Foundation, Yale University*, 1:1–7, 1962.
- [Pat68] PATTERSON, TNL: The Optimum addition of points to quadrature formulae. *Math. Comput*, 22:847–856, 1968.
- [Pre03] PREKOPA, ANDRAIS: Probabilistic Programming. 10:267 – 351, 2003.
- [RWM03] RIPHAHN, R.T., A. WAMBACH und A. MILLION: Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of applied econometrics*, 18(4):387–405, 2003.
- [RZ07] RICHARD, J.F. und W. ZHANG: Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411, 2007.
- [Sch84] SCHERVISH, M.J.: Algorithm AS 195: Multivariate normal probabilities with error bound. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(1):81–94, 1984.
- [Sha88] SHAW, J. E. H.: A Quasirandom Approach to Integration in Bayesian Statistics. *The Annals of Statistics*, 16(2):pp. 895–914, 1988.
- [Sob76] SOBOL', I.M.: Uniformly distributed sequences with an additional property of uniformity. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 16(5):1332–1337, 1976.
- [Ste92] STERN, S.: A method for smoothing simulated moments of discrete probabilities in multinomial probit models. *Econometrica*, 60(4):943–952, 1992.
- [Sto94] STOER, J.: Numerische Mathematik 1 und 2. *Springer-Verlag, Berlin*, 1994.
- [Sz76] SZANTAI, T.: An algorithm for calculating values and gradient vectors for the multiple normal distribution function (in Hungarian). *Alkalmazott Matematikai Lapok*, 2:27–39, 1976.
- [Tra03] TRAIN, K.: Discrete choice methods with simulation. *Cambridge Univ Pr*, 2003.
- [Woo02] WOOLDRIDGE, J.M.: Econometric analysis of cross section and panel data. *The MIT press*, 2002.

