

Simulationsbasierter Vergleich von Homogenitätstests

Bachelorarbeit
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

vorgelegt von
Sabrina Thorn

September 2010
Betreuer: Prof. Dr. H. Holzmann

Inhaltsverzeichnis

Einleitung	1
1 Mischungsmodelle	4
1.1 Endliche Mischungsmodelle	4
1.2 Identifizierbarkeit in Mischungsmodellen	6
2 Testen auf Homogenität in Mischungsmodellen	10
2.1 Testen auf Homogenität mit Likelihood-basierten Tests	10
2.1.1 Der Likelihood-Quotienten-Test	10
2.1.2 Der modifizierte Likelihood-Quotienten-Test	32
2.2 Der Bootstrap-Test	34
3 Simulationsstudie	36
3.1 Likelihood-Quotienten-Test	36
3.2 Vergleich der Testniveaus des MLQT und des Bootstrap-Tests	42
3.3 Vergleich der Power des MLQT und des Bootstrap-Tests	45
4 Ausblick	51
A Die Maximum-Likelihood-Methode	52
B Der Likelihood-Quotienten-Test	54
C Sätze zur Identifizierbarkeit von endlichen Mischungen	56
D Wahrscheinlichkeits- und momenterzeugende Funktionen	58
Literaturverzeichnis	60
Eigenständigkeitserklärung	62

Einleitung

Diese Arbeit befasst sich mit zwei statistischen Tests, mit denen sich eine Population auf Homogenität überprüfen lässt. Eine Population ist in Bezug auf ein zu betrachtendes Merkmal ihrer Mitglieder homogen, wenn die Wahrscheinlichkeitsverteilung des Merkmals für jedes Mitglied identisch ist. Lässt sich jedoch die Population in mehrere Subpopulationen unterteilen, zwischen denen sich die Verteilung des Merkmals unterscheidet, so spricht man von einer heterogenen Population. Dabei gehören die Verteilungen meistens zur gleichen parametrischen Familie von Wahrscheinlichkeitsverteilungen, wobei $\{f(x; \theta); \theta \in \Theta \subseteq \mathbb{R}^d, d \geq 1\}$ die zugehörigen Dichten bezüglich eines σ -endlichen Maßes sind und sich lediglich der Parameter θ zwischen den Subpopulationen unterscheidet.

Häufig sind bei der Untersuchung einer Stichprobe lediglich die Ausprägungen des Merkmals beobachtbar, ohne dass festzustellen ist, aus welchen Subpopulationen diese Daten stammen. In diesem Fall beschreibt man die Verteilung des Merkmals durch eine Mischung aus den Verteilungen der einzelnen Subpopulationen, wobei diese Verteilungen dem Anteil der Subpopulationen an der gesamten Population entsprechend gewichtet werden.

Die Arbeit beschränkt sich auf Mischungen aus zwei Verteilungen aus jeweils denselben parametrischen Verteilungsfamilien. Der Parameter θ ist dabei stets univariat. Die Mischungsdichten sind im Folgenden also immer von der Form

$$(1 - \alpha)f(x; \theta_1) + \alpha f(x; \theta_2),$$

wobei α und $1 - \alpha$ die Anteile der Subpopulationen an der gesamten Population sind und $f(x; \theta_i)$ mit $\theta_i \in \mathbb{R}$ ($i = 1, 2$) die Wahrscheinlichkeitsdichte des Merkmals in der jeweiligen Subpopulation ist. Das bedeutet, dass man bei einer Stichprobe mit der Wahrscheinlichkeit $1 - \alpha$ ein Mitglied der Subpopulation wählt, in der das Merkmal die Dichte $f(x; \theta_1)$ besitzt, und mit der Wahrscheinlichkeit α wählt man ein Mitglied aus der anderen Subpopulation,

in der das Merkmal die Dichte $f(x; \theta_2)$ besitzt.

Der Hauptteil der Arbeit besteht im Vergleich zweier Homogenitätstests – einer modifizierten Version des Likelihood-Quotienten-Tests und des Bootstrap-Tests. Mit diesen lässt sich die Hypothese „Die Population ist homogen“ gegen die Alternative „Die Population ist heterogen“ testen.

Eine wichtige Voraussetzung für die Durchführung der beiden Tests ist die Identifizierbarkeit der Mischungsverteilungen. Eine parametrische Familie von Wahrscheinlichkeitsverteilungen ist identifizierbar, wenn verschiedene Parameterwerte verschiedene Verteilungen aus der Familie bestimmen. Endliche Mischungsverteilungen sind in der Regel jedoch aufgrund ihrer Invarianz bezüglich der Komponentenreihenfolge und aufgrund möglichen Overfittings nicht identifizierbar. Ihre Identifizierbarkeit kann jedoch durch Einschränkungen der Parameter und Gewichte erreicht werden.

Es wird außerdem auf die asymptotische Verteilung der Teststatistik unter der Hypothese eingegangen. In endlichen Mischungsmodellen handelt es sich bei der asymptotischen Nullverteilung der Likelihood-Quotienten-Test-Statistik nicht mehr um eine gewöhnliche χ^2 -Verteilung, da dafür notwendige Regularitätsbedingungen nicht erfüllt werden, sondern um das Supremum eines Gaußprozesses, welches nur schwer zu bestimmen und für weiterführende Rechnungen nicht gut geeignet ist. Eine einfachere asymptotische Nullverteilung erhält man für die Teststatistik des modifizierten Likelihood-Quotienten-Tests von Chen et al. (2001). Dabei handelt es sich um eine Mischung aus χ^2 -Verteilungen.

Die Arbeit beginnt in Kapitel 1 mit einer Einführung in endliche Mischungsmodelle, da diese die Ausgangslage für die Untersuchung der Homogenitätstests bilden. Es wird gezeigt, warum endliche Mischungsverteilungen im Allgemeinen nicht identifizierbar sind und welche Einschränkungen vorgenommen werden müssen, um Identifizierbarkeit in Mischungsmodellen herzustellen. In Kapitel 2 werden zunächst der Likelihood-Quotienten-Test sowie der von Chen et al. (2001) entwickelte modifizierte Likelihood-Quotienten-Test in Mischungsmodellen vorgestellt und die asymptotische Nullverteilung ihrer Teststatistiken betrachtet. Anschließend wird der Bootstrap-Test erläutert, um ihn dann in einer Simulationsstudie mit dem modifizierten Likelihood-Quotienten-Test zu vergleichen. Außerdem wird die Abhängigkeit der Likelihood-Quotienten-Test-Statistik vom Parameter θ verdeutlicht. Die Vorgehensweisen und Beobachtungen dieser Studie werden in Kapitel 3 dargestellt und in Kapitel 4 ausgewertet. Im Anhang werden die Maximum-Likelihood-Methode und der Likelihood-Quotienten-Test in regulären Model-

len beschrieben und für die in Kapitel 2 durchgeführten Beweise zur Identifizierbarkeit von endlichen Mischungsverteilungen relevante Sätze und Definitionen aufgeführt.

Kapitel 1

Mischungsmodelle

1.1 Endliche Mischungsmodelle

In der Statistik finden Mischungsmodelle häufig Anwendung in der Modellierung von Daten heterogener Populationen. Eine solche Population lässt sich in m Subpopulationen aufspalten, zwischen denen sich die Wahrscheinlichkeitsverteilung des zu betrachtenden Merkmals X , das als diskrete oder stetige Zufallsvariable aufgefasst werden kann, unterscheidet. Meistens stammen die Verteilungen aus derselben parametrischen Familie von Wahrscheinlichkeitsverteilungen mit zugehörigen Dichten $\{f(x; \theta); \theta \in \Theta \subseteq \mathbb{R}^d, d \geq 1\}$ bezüglich eines dominierenden σ -endlichen Maßes ν , wobei sich lediglich der Parameter θ zwischen den Subpopulationen unterscheidet. Da die Ausprägung von X von bestimmten Faktoren der Subpopulationen abhängt, ist bei der Untersuchung einer Stichprobe stets zu berücksichtigen, aus welchen Subpopulationen die Daten stammen. Mithilfe einer diskreten Zufallsvariablen S mit Werten in $\{1, \dots, m\}$ kann man für eine aus der Population gezogene Stichprobe X_1, \dots, X_n die Beobachtungen den jeweiligen Subpopulationen zuordnen. Allerdings lassen sich häufig nur die Ausprägungen x_1, \dots, x_n von X beobachten, ohne aber feststellen zu können, aus welchen Subpopulationen diese Informationen stammen. Für eine Modellierung, bei der diese Umstände berücksichtigt werden, verwendet man Mischungsmodelle, wie sie u. a. Frühwirth-Schnatter (2006), Schlattmann (2009) oder McLachlan und Peel (2000) in ihren Werken untersucht haben.

Die gemeinsame Dichte von X und S bezüglich $\nu \times (\text{Zählmaß})$ ist gegeben

durch

$$p(x, s; \Psi^{(m)}) = p(x; \Psi^{(m)} | s) p(s) = f(x; \theta_s) \alpha_s,$$

wobei $f(x; \theta_s)$ die Dichte von X bezüglich ν in Subpopulation s ist und mit $0 \leq \alpha_s \leq 1$ gewichtet wird. Somit ist die Randdichte von X gegeben durch die Mischungsichte

$$f_{mix}(x; \Psi^{(m)}) = \sum_{s=1}^m p(x, s; \Psi^{(m)}) = \sum_{s=1}^m f(x; \theta_s) \alpha_s.$$

Dabei ist

$$\Psi^{(m)} = (\alpha_1, \dots, \alpha_{m-1}, \theta_1, \dots, \theta_m)$$

ein Vektor aus dem Parameterraum

$$[0, 1]^{m-1} \times \Theta^m =: \Theta_m.$$

Der Index m steht für die Anzahl der Komponenten der Mischungsichte und es gilt

$$\alpha_m = 1 - \sum_{i=1}^{m-1} \alpha_i.$$

Die Gewichte $\alpha_1, \dots, \alpha_m$ werden als Anteile der jeweiligen Subpopulationen an der gesamten zu betrachtenden Population interpretiert. Die Wahrscheinlichkeit bei einer Stichprobe die Subpopulationen i zu wählen ist

$$P(S = i) = \alpha_i \quad \text{mit} \quad i = 1, \dots, m.$$

Stammen die Dichten der Komponenten aus der gleichen Verteilungsfamilie, so wird f als *Kernfunktion* bezeichnet. Zusammenfassend gilt nach Frühwirth-Schnatter (2006):

Definition 1.1.1 Eine reelle Zufallsvariable X besitzt eine endliche Mischungsverteilung, wenn die zugehörige Dichtefunktion für alle $x \in \mathbb{R}$ von der Form

$$f_{mix}(x; \Psi^{(m)}) = \alpha_1 f(x; \theta_1) + \dots + \alpha_m f(x; \theta_m)$$

ist, wobei $f(x; \theta_i)$ ($i = 1, \dots, m$) die Dichtefunktionen von X bezüglich eines dominierenden, σ -endlichen Maßes ν in den entsprechenden Subpopulationen sind, m die Anzahl der Komponenten ist und für die Gewichte $\alpha_i \geq 0$ für alle $i = 1, \dots, m$ und $\sum_{i=1}^m \alpha_i = 1$ gilt.

Die Arbeit beschränkt sich auf einparametrische Mischungsmodelle mit zwei Komponenten. Mit dem modifizierten Likelihood-Quotienten-Test von Chen et al. (2001) und dem Bootstrap-Test wird beim Testen auf Homogenität also überprüft, ob eine Mischung aus einer Komponente oder aus zwei Komponenten besteht.

1.2 Identifizierbarkeit in Mischungsmodellen

Beim modifizierten Likelihood-Quotienten-Test sowie beim Bootstrap-Test werden Schätzer für die Parameter der Mischungsdichte mithilfe der Maximum-Likelihood-Methode¹ ermittelt. Dabei müssen Mischungsdichten identifizierbar sein, um eindeutige Parameter zu erhalten. Im Allgemeinen ist eine parametrische Familie von Wahrscheinlichkeitsdichten identifizierbar, wenn verschiedene Parameterwerte verschiedene Dichten aus der Familie generieren.

Definition 1.2.1 Die parametrische Familie von Wahrscheinlichkeitsdichten $\{f(x; \Psi); \Psi \in \Theta\}$ ist genau dann identifizierbar, wenn die Gleichheit der Dichten

$$f(x; \Psi) = f(x; \Psi^*) \quad \text{für fast alle } x$$

die Gleichheit der Parameter

$$\Psi = \Psi^*$$

impliziert (vgl. Frühwirth-Schnatter (2006)).

In Mischungsmodellen sind Dichten in der Regel jedoch nicht identifizierbar. Zwei Gründe dafür sind die Invarianz der Mischungsdichten bezüglich der Reihenfolge der Komponenten und mögliches Overfitting. Da sich die Arbeit auf einparametrische Mischungsmodelle mit zwei Komponenten beschränkt, wird im Folgenden lediglich dieser Fall behandelt. Für weitere Informationen siehe Frühwirth-Schnatter (2006).

Betrachtet man zunächst die Mischungsdichte

$$f_{mix}(x; \Psi^{(2)}) = (1 - \alpha)f(x; \theta_1) + \alpha f(x; \theta_2) \tag{1.1}$$

¹siehe Anhang A

mit $\Psi^{(2)} = (\alpha, \theta_1, \theta_2)$, so erhält man durch das Vertauschen der Komponenten den Parametervektor $\Psi^{*(2)} = (1 - \alpha, \theta_2, \theta_1)$, der sich von $\Psi^{(2)}$ unterscheidet. Allerdings werden durch $\Psi^{(2)}$ und $\Psi^{*(2)}$ dieselben Mischungsdichten bestimmt:

$$\begin{aligned} f_{mix}(x; \Psi^{*(2)}) &= \alpha f(x; \theta_2) + (1 - \alpha) f(x; \theta_1) \\ &= (1 - \alpha) f(x; \theta_1) + \alpha f(x; \theta_2) = f_{mix}(x; \Psi^{(2)}). \end{aligned}$$

Die Familie der Dichten, aus der $f_{mix}(x; \Psi^{(2)})$ bzw. $f_{mix}(x; \Psi^{*(2)})$ stammen, ist somit nicht identifizierbar.

Neben der Invarianz der Mischungsverteilung bezüglich der Komponentenanzahl kann auch Overfitting die Ursache für Nichtidentifizierbarkeit sein. Hierbei setzen sich Mischungsdichten aus mehr Komponenten zusammen, als zu ihrer Darstellung nötig wären. Dabei lassen sich zwei Fälle unterscheiden. Zum einen kann jede Mischungsverteilung mit $m - 1$ Komponenten als Mischungsverteilung mit m Komponenten dargestellt werden, indem die m -te Komponente mit Gewicht $\alpha_m = 0$ angefügt wird. Man spricht hier auch von einer leeren Komponente. Für die Mischungsdichte 1.1 gilt

$$\begin{aligned} f_{mix}(x; \Psi^{(2)}) &= (1 - \alpha) f(x; \theta_1) + \alpha f(x; \theta_2) \\ &= (1 - \alpha) f(x; \theta_1) + \alpha f(x; \theta_2) + 0 \cdot f(x; \theta_3) = f_{mix}(x; \Psi^{(3)}) \end{aligned}$$

mit $\Psi^{(3)} = (\alpha, 0, \theta_1, \theta_2, \theta_3)$. Solange Gewichte mit Wert 0 zugelassen werden, sind Mischungsdichten also nicht identifizierbar.

Der zweite Fall von Overfitting tritt auf, wenn eine Komponente einer Mischungsverteilung in zwei Komponenten aufgespalten wird oder, äquivalent, die Dichten zweier Komponenten identisch sind. Für die Mischungsdichte 1.1 bedeutet dies

$$\begin{aligned} f_{mix}(x; \Psi^{(2)}) &= (1 - \alpha) f(x; \theta_1) + \alpha f(x; \theta_2) \\ &= (1 - \alpha) f(x; \theta_1) + (\alpha - \tilde{\alpha}) f(x; \theta_2) + \tilde{\alpha} f(x; \theta_2) \\ &= f_{mix}(x; \Psi^{(3)}) \end{aligned}$$

mit $\Psi^{(3)} = (\alpha - \tilde{\alpha}, \tilde{\alpha}, \theta_1, \theta_2, \theta_3)$ und $0 \leq \tilde{\alpha} \leq \alpha$.

Diese beiden Arten des Overfittings lassen sich verallgemeinern: Es existiert zu jeder Mischungsdichte mit zwei Komponenten eine äquivalente Mischungsdichte mit $2 + l$ ($l \in \mathbb{N}$) Komponenten, von denen entweder l Komponenten leer sind oder $l + 1$ Komponenten die gleiche Dichte besitzen. Die obigen

Ausführungen beziehen sich lediglich auf den Fall $l = 1$, lassen sich jedoch leicht auf andere Fälle übertragen.

Durch die Einführung einfacher Restriktionen der Parameter kann die Identifizierbarkeit einer Familie von Mischungsdichten hergestellt werden. Mithilfe einer Regel, die die Reihenfolge der Komponenten bestimmt, lässt sich das Problem der beliebigen Anordnung der Komponenten umgehen. So implizieren die Bedingungen

$$\theta_1 < \dots < \theta_m$$

oder

$$\alpha_1 < \dots < \alpha_m$$

jeweils eine eindeutige Reihenfolge der Komponenten. Dabei wird jedoch noch nicht die Nichtidentifizierbarkeit aufgrund von Overfitting berücksichtigt. Hierfür müssen die Parameter weiter eingeschränkt werden. Die Bedingung

$$\alpha_i > 0 \quad \text{für alle } i = 1, \dots, m$$

lässt keine mit 0 gewichteten Komponenten zu und schließt somit Nichtidentifizierbarkeit aufgrund leerer Komponenten aus. Die Bedingung

$$\theta_i \neq \theta_j \quad \text{für alle } i, j = 1, \dots, m, \quad i \neq j$$

verhindert, dass mehrere Komponenten die gleiche Dichte besitzen, sodass Nichtidentifizierbarkeit aufgrund gleicher Komponentendichten ebenfalls vermieden wird.

Die Definition 1.2.1 von Identifizierbarkeit muss für Mischungsmodelle also entsprechend angepasst werden:

Definition 1.2.2 Seien

$$f(x; \Psi) = \sum_{i=1}^m \alpha_i f(x, \theta_i) \quad \text{und} \quad f(x; \Psi^*) = \sum_{i=1}^{m^*} \alpha_i^* f(x, \theta_i^*)$$

Mischungsdichten aus der Familie $\{f(x; \Psi); \Psi \in \Theta\}$. Die Gewichte α_i und α_i^* seien für alle i strikt positiv und für jede Mischungsdichte seien alle Komponentenparameter verschieden. Dann ist die Familie der Mischungsdichten genau dann identifizierbar, wenn die Gleichheit der Dichten

$$f(x; \Psi) = f(x; \Psi^*) \quad \text{für fast alle } x$$

impliziert, dass

$$m = m^*$$

gilt und die Dichten bis auf Permutation der Komponenten gleich sind, d. h., dass nach geeigneter Anordnung der Komponenten

$$\alpha_i = \alpha_i^* \quad \text{und} \quad f(x; \theta_i) = f(x; \theta_i^*) \quad \text{für} \quad i = 1, \dots, m$$

gilt (vgl. Yakowitz und Spragins (1968)).

Kapitel 2

Testen auf Homogenität in Mischungsmodellen

2.1 Testen auf Homogenität mit Likelihood-basierten Tests

2.1.1 Der Likelihood-Quotienten-Test

Sei X_1, \dots, X_n eine unabhängige Stichprobe der Mischungsdichte

$$f_{mix}(x; \Psi^{(2)}) = (1 - \alpha)f(x; \theta_1) + \alpha f(x; \theta_2), \quad (2.1)$$

mit $\Psi^{(2)} = (\alpha, \theta_1, \theta_2) \in [0, 1] \times \Theta^2$.

Um die zugrundeliegende Population auf Homogenität zu untersuchen, wird die Hypothese

$$H_0 : \alpha(1 - \alpha)(\theta_1 - \theta_2) = 0$$

gegen die Alternative

$$H_1 : \alpha(1 - \alpha)(\theta_1 - \theta_2) \neq 0$$

getestet. Die Hypothese ist genau dann erfüllt, wenn $\alpha \in \{0, 1\}$ oder $\theta_1 = \theta_2$ gilt. Es wird also überprüft, ob sich die Mischungsdichte aus einer Komponente oder aus zwei verschiedenen Komponenten zusammensetzt. Mithilfe der Log-Likelihood-Funktion

$$l_n(\alpha, \theta_1, \theta_2) = \sum_{i=1}^n \log f_{mix}(X_i; \Psi^{(2)})$$

lassen sich Schätzer für die Parameter unter der Alternative und der Hypothese berechnen. Seien

$$(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2) = \operatorname{argmax}_{[0,1] \times \Theta^2} l_n(\alpha, \theta_1, \theta_2) \quad \text{und} \quad \hat{\theta}_0 = \operatorname{argmax}_{\Theta} l_n(1/2, \theta, \theta)$$

die jeweiligen Maximum-Likelihood-Schätzer. Die Likelihood-Quotienten-Test-Statistik (LQT-Statistik) wird definiert als

$$R_n = 2 \left\{ l_n(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2) - l_n(1/2, \hat{\theta}_0, \hat{\theta}_0) \right\}$$

und die Nullhypothese wird verworfen, wenn R_n groß ist.

Unter den in Anhang B aufgezeigten Regularitätsbedingungen ist nach Ferguson (1996) die asymptotische Nullverteilung einer LQT-Statistik eine χ^2 -Verteilung. Allerdings werden im betrachteten Mischungsmodell 2.1 zwei dieser Bedingungen nicht erfüllt. Zum einen liegt die Nullhypothese für $\alpha = 0$ oder $\alpha = 1$ auf dem Rand des Parameterraums $\Theta_2 = [0, 1] \times \Theta^2$, zum anderen ist die Mischungsdichte unter der Nullhypothese nicht identifizierbar. So impliziert beispielsweise die Gleichheit von $\frac{1}{2}f(x; \theta) + \frac{1}{2}f(x; \theta)$ und $f(x; \theta) + 0 \cdot f(x; \theta')$ nicht die Gleichheit der Parameter.

Zur Berechnung der asymptotischen Verteilung der Teststatistik R_n unter der Hypothese stellen Chen et al. (2001) Regularitätsbedingungen an die Kernfunktion, die von den von Ferguson (1996) angegebenen Bedingungen abweichen. Dafür werden zunächst folgende Zufallsvariablen definiert:

$$\begin{aligned} Y_i(\theta) &= \frac{f(X_i; \theta) - f(X_i; \theta_0)}{(\theta - \theta_0)f(X_i; \theta_0)}, & \theta \neq \theta_0 \\ Y_i &= Y_i(\theta_0) = \frac{f'(X_i; \theta_0)}{f(X_i; \theta_0)} \\ Z_i(\theta) &= \frac{Y_i(\theta) - Y_i(\theta_0)}{\theta - \theta_0}, & \theta \neq \theta_0 \\ Z_i &= Z_i(\theta_0) = \frac{f''(X_i; \theta_0)}{2f(X_i; \theta_0)}. \end{aligned} \tag{2.2}$$

Außerdem sei

$$h(\theta) = \frac{E[Y_i(\theta)Z_i(\theta)]}{E[Y_i^2]}.$$

Die Regularitätsbedingungen nach Chen et al. (2001) lauten nun:

A1. $\Theta \subset \mathbb{R}$ ist kompakt.

A2. Für jedes $\theta \in \Theta$ gilt:

(a) $E|\log f(X; \theta)| < \infty$,

(b) es existiert ein $\rho > 0$, sodass $E[\log \{f(X, \theta, \rho)\}] < \infty$ für
 $f(x, \theta, \rho) = 1 + \sup_{|\theta' - \theta| \leq \rho} \{f(x; \theta')\}$.

(Wald'sche Integrationsbedingungen)

A3. Der Träger der Kernfunktion $f(x; \theta)$ ist unabhängig von θ und $f(x; \theta)$ ist bezüglich θ zweimal stetig differenzierbar. (Glattheitsbedingungen)

A4. Die Kernfunktion $f(x; \theta)$ ist stark identifizierbar, d. h.

(a) $f_{mix}(x; \Psi^{(2)}) = f_{mix}(x; \Psi^{*(2)})$ impliziert $\Psi^{(2)} = \Psi^{*(2)}$ (die Gleichheit gilt bis auf Permutation der Komponenten),

(b) für alle $\theta_1 \neq \theta_2$ aus Θ und für alle x gilt:

$$\sum_{j=1}^2 \{a_j f(x; \theta_j) + b_j f'(x; \theta_j) + c_j f''(x; \theta_j)\} = 0$$

impliziert $a_j = b_j = c_j = 0$ mit $j = 1, 2$.

(Starke Identifizierbarkeit)

A5. Es existiert eine Funktion

$g : \mathbb{R} \rightarrow \mathbb{R}$ mit $E[g(X)] < \infty$, sodass gilt:

(a) $|Y_i(\theta)|^3 \leq g(X_i)$ und $|Z_i(\theta)|^3 \leq g(X_i)$ für alle $\theta \in \Theta$,

(b) $|Z_i''(\theta)|^2 \leq g(X_i)$ für θ aus einer Umgebung von θ_0 .

(Starkes Gesetz der großen Zahlen)

Satz 2.1.1 (Chen et al., 2001) *Unter den Regularitätsbedingungen A1-A5 und der Hypothese gilt*

$$R_n \xrightarrow{d} \sup_{\theta \in \Theta} \{W^+(\theta)\}^2.$$

Hierbei bezeichne $W(\theta)$ einen Gaußprozess mit Erwartungswert 0, Varianz 1 und Autokorrelationsfunktion

$$\rho(\theta, \theta') = \frac{\text{cov} \{Z_i(\theta) - h(\theta)Y_i, Z_i(\theta') - h(\theta')Y_i\}}{\sqrt{\text{var} \{Z_i(\theta) - h(\theta)Y_i\} \text{var} \{Z_i(\theta') - h(\theta')Y_i\}}}.$$

Anmerkung: $R_n \xrightarrow{d} \sup_{\theta \in \Theta} \{W^+(\theta)\}^2$ bedeutet, dass R_n in Verteilung gegen eine Zufallsvariable X konvergiert, deren Verteilung jener Gaussprozess ist. Somit gilt $\int f dP_{R_n} \xrightarrow{n \rightarrow \infty} \int f dP_X$ für alle stetigen und beschränkten Funktionen f .

In der Simulationsstudie (Kapitel 3) werden die Homogenitätstests anhand von Mischungen von Poisson-, Normal- und Binomialverteilungen untersucht. Vorher ist aber noch zu zeigen, dass diese Kernfunktionen die Regularitätsbedingungen erfüllen. Im Folgenden werden die verschiedenen Funktionen auf einige der Bedingungen überprüft.

Beispiel 2.1.2 *Für poissonverteilte Zufallsvariablen X_i mit Dichte*

$$f(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

erhält man für die Zufallsvariablen 2.2

$$\begin{aligned} Y_i(\lambda) &= \frac{\lambda^{X_i} \lambda_0^{-X_i} e^{\lambda_0 - \lambda} - 1}{\lambda - \lambda_0} \\ Y_i(\lambda_0) &= \frac{X_i}{\lambda_0} - 1 \\ Z_i(\lambda) &= \frac{\lambda^{X_i} \lambda_0^{-X_i} e^{\lambda_0 - \lambda} - 1}{(\lambda - \lambda_0)^2} - \frac{X_i - \lambda_0}{\lambda_0(\lambda - \lambda_0)} \\ Z_i(\lambda_0) &= \frac{(X_i - \lambda_0)^2 - X_i}{2\lambda_0^2}. \end{aligned}$$

A1. Der Parameterraum $\Theta = [c, d]$ mit $0 < c < d$ und $c, d \in \mathbb{R}$ ist kompakt.

A2. (a) Es wird lediglich der Fall $|\lambda| < 1$ betrachtet. Hier gilt

$$\begin{aligned} & E \left| \log \left(\frac{\lambda^X}{X!} e^{-\lambda} \right) \right| \\ &= E \left| \log \lambda^X + \log e^{-\lambda} + (-\log X!) \right| \\ &\leq E |\log \lambda^X| + E |\log e^{-\lambda}| + E |-\log X!| \\ &= E |X \log \lambda| + \lambda + E |\log X!|. \end{aligned}$$

Da λ aus dem kompakten Parameterraum $\Theta = [c, d]$ stammt, sind λ und $E |X \log \lambda| = |\log \lambda| |\lambda|$ endlich. Mit der Jensenschen Ungleichung lässt sich $E |\log X!|$ durch $\log EX!$ nach oben abschätzen. Es bleibt somit noch zu zeigen, dass $EX!$ endlich ist. Der Erwartungswert von $X!$ ist

$$EX! = \sum_{k \geq 0} k! \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k \geq 0} \lambda^k e^{-\lambda}.$$

Mithilfe des Quotientenkriteriums lässt sich zeigen, dass die Reihe $\sum_{k \geq 0} \lambda^k e^{-\lambda}$ konvergiert:

$$\left| \frac{\lambda^{k+1} e^{-\lambda}}{\lambda^k e^{-\lambda}} \right| = |\lambda| < 1.$$

Also ist $EX!$ endlich und es gilt insgesamt

$$E \left| \log \left(\frac{\lambda^X}{X!} e^{-\lambda} \right) \right| < \infty.$$

A3. Der Träger der Poissonverteilung ist \mathbb{N}_0 . Dieser ist unabhängig von λ . Weiterhin ist $f(x; \lambda)$ zweimal stetig differenzierbar bezüglich λ :

$$\begin{aligned} f'(x; \lambda) &= \frac{e^{-\lambda} \lambda^{x-1} (x - \lambda)}{x!} \\ f''(x; \lambda) &= \frac{e^{-\lambda} \lambda^{x-2} ((x - \lambda)^2 - x)}{x!}. \end{aligned}$$

- A4. (a) Es sei $\mathcal{F} = \{F_\lambda : F_\lambda = F(x, \lambda)\}$ eine Familie von Poissonverteilungen mit Parameter λ . Für den Beweis der Identifizierbarkeit von endlichen Mischungen von \mathcal{F} wird Satz C.1 von Teicher (1963) verwendet. \mathcal{F} lässt sich durch

$$F_{\lambda_1} = F(x; \lambda_1) \prec F(x; \lambda_2) = F_{\lambda_2},$$

falls $\lambda_1 > \lambda_2$, ordnen. Die wahrscheinlichkeitserzeugende Funktion der Poissonverteilung

$$\phi(t) = e^{\lambda(t-1)}$$

ist bijektiv und linear. Für zwei beliebige Parameter λ_1 und λ_2 besitzen die entsprechenden Funktionen ϕ_1 und ϕ_2 denselben Definitionsbereich $S_\phi = (-\infty, \infty)$. Insbesondere gilt also $S_{\phi_1} \subseteq S_{\phi_2}$. Weiterhin gilt für $t_1 = +\infty$

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow +\infty} \frac{e^{\lambda_2(t-1)}}{e^{\lambda_1(t-1)}} = \lim_{t \rightarrow +\infty} e^{(\lambda_2 - \lambda_1)(t-1)} = 0,$$

da $\lambda_1 > \lambda_2$ ist. Somit folgt nach Satz C.1 die Identifizierbarkeit von endlichen Mischungen von Poissonverteilungen.

- (b) Der Beweis zur starken Identifizierbarkeit von endlichen Mischungen von Poissonverteilungen findet sich bei Chen (1995). Sei

$$\sum_{j=1}^2 \{a_j f(x; \lambda_j) + b_j f'(x; \lambda_j) + c_j f''(x; \lambda_j)\} = 0. \quad (2.3)$$

Es ist zu zeigen, dass hieraus $a_j = b_j = c_j = 0$ folgt. Durch die momenterzeugende Funktion der Poissonverteilung

$$m(t) = \sum_{n=0}^{\infty} e^{tn} f(n; \lambda) = e^{\lambda(e^t - 1)}$$

erhält man

$$\begin{aligned} & \sum_{n=0}^{\infty} e^{tn} \left(\sum_{j=1}^2 \{a_j f(n; \lambda_j) + b_j f'(n; \lambda_j) + c_j f''(n; \lambda_j)\} \right) \\ &= \sum_{j=1}^2 \left(a_j \sum_{n=0}^{\infty} e^{tn} f(n; \lambda_j) + b_j \sum_{n=0}^{\infty} e^{tn} f'(n; \lambda_j) + c_j \sum_{n=0}^{\infty} e^{tn} f''(n; \lambda_j) \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^2 \left(a_j \sum_{n=0}^{\infty} e^{tn} \frac{\lambda_j^n e^{-\lambda_j}}{n!} + b_j \sum_{n=0}^{\infty} e^{tn} \frac{n\lambda_j^{n-1} e^{-\lambda_j} - \lambda_j^n e^{-\lambda_j}}{n!} + \right. \\
&\quad \left. + c_j \sum_{n=0}^{\infty} e^{tn} \frac{\lambda_j^{n-2} e^{-\lambda_j} ((n - \lambda_j)^2 - n)}{n!} \right) \\
&= \sum_{j=1}^2 (a_j e^{\lambda_j(e^t-1)} + b_j e^{\lambda_j(e^t-1)} e^t - b_j e^{\lambda_j(e^t-1)} + c_j e^{\lambda_j(e^t-1)} \\
&\quad - 2c_j e^{\lambda_j(e^t-1)} e^t + c_j e^{\lambda_j(e^t-1)} e^{2t}) \\
&= \sum_{j=1}^2 [(a_j - b_j + c_j) + (b_j - 2c_j)e^t + c_j e^{2t}] e^{\lambda_j(e^t-1)}.
\end{aligned}$$

Es soll gelten, dass

$$\sum_{j=1}^2 [(a_j - b_j + c_j) + (b_j - 2c_j)e^t + c_j e^{2t}] e^{\lambda_j(e^t-1)} = 0$$

ist. Sei $\lambda_{max} = \max\{\lambda_1, \lambda_2\}$. Dann strebt $e^{\lambda_{max}(e^t-1)+2t}$ für $t \rightarrow \infty$ am schnellsten gegen ∞ . Also muss $c_j = 0$ gelten. Nun strebt $e^{\lambda_j(e^t-1)+t}$ am schnellsten gegen ∞ , sodass $b_j = 0$ sein muss. Schließlich muss auch $a_j = 0$ gelten, damit die Gleichung erfüllt ist. Somit ist $a_j = b_j = c_j = 0$ die einzige Lösung für 2.3 und es folgt, dass endliche Mischungen von Poissonverteilungen stark identifizierbar sind.

A5.

$$\begin{aligned}
|Y_i(\lambda)|^3 &= \left| \frac{f(X_i; \lambda) - f(X_i; \lambda_0)}{(\lambda - \lambda_0)f(X_i; \lambda_0)} \right|^3 \\
&= \left| \frac{f'(X_i; \tilde{\lambda})}{f(X_i; \lambda_0)} \right|^3 \quad \text{mit } \tilde{\lambda} \in (\lambda_0, \lambda) \text{ für o.B.d.A. } \lambda_0 < \lambda \\
&= \left| \frac{e^{-\tilde{\lambda}} \tilde{\lambda}^{X_i-1} (X_i - \tilde{\lambda})}{\lambda_0^{X_i} e^{-\lambda_0}} \right|^3 \\
&= \left| e^{\lambda_0 - \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{X_i} \cdot \frac{1}{\tilde{\lambda}} \cdot (X_i - \tilde{\lambda}) \right|^3 \\
&= \left| e^{3(\lambda_0 - \tilde{\lambda})} \cdot \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{3X_i} \cdot \frac{1}{\tilde{\lambda}^3} \cdot (X_i - \tilde{\lambda})^3 \right| \\
&= \left| e^{3(\lambda_0 - \tilde{\lambda})} \cdot \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{3X_i} \cdot \frac{1}{\tilde{\lambda}^3} \cdot (X_i^3 - 3X_i^2\tilde{\lambda} + 3X_i\tilde{\lambda}^2 - \tilde{\lambda}^3) \right| \\
&\leq \left| e^{3(\lambda_0 - c)} \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot \frac{1}{c^3} \cdot (X_i^3 + 3X_i^2d + 3X_id^2 + d^3) \right| \quad \text{da } \tilde{\lambda} \in (\lambda_0, \lambda) \subset [c, d] \\
&= \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot \frac{1}{c^3} \cdot (X_i^3 + 3X_i^2d + 3X_id^2 + d^3) \right| \quad \text{mit } a := e^{3(\lambda_0 - c)} \\
&\leq \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot \frac{1}{c^3} \cdot (3!e^{X_i} + 3 \cdot 2!c e^{X_i} + 3c^2 e^{X_i} + c^3 e^{X_i}) \right| \\
&= \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot a_1 e^{X_i} \right| \quad \text{mit } a_1 := \frac{3! + 3 \cdot 2!c + 3c^2 + c^3}{c^3} \\
&= \left| aa_1 \cdot e^{X_i \left(3 \log\left(\frac{d}{\lambda_0}\right) + 1 \right)} \right|
\end{aligned}$$

$$\begin{aligned}
|Z_i(\lambda)|^3 &= \left| \frac{Y_i(\lambda) - Y_i(\lambda_0)}{\lambda - \lambda_0} \right|^3 \\
&= \left| Y_i'(\tilde{\lambda}) \right|^3 \quad \text{mit } \tilde{\lambda} \in (\lambda_0, \lambda) \text{ für o.B.d.A. } \lambda_0 \leq \lambda \\
&= \left| \frac{f''(X_i; \tilde{\lambda})}{f(X_i; \lambda_0)} \right|^3 \\
&= \left| \frac{e^{-\tilde{\lambda}} \tilde{\lambda}^{X_i-2} \left((X_i - \tilde{\lambda})^2 - X_i \right)}{e^{-\lambda_0} \lambda_0^{X_i}} \right|^3 \\
&= \left| e^{\lambda_0 - \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{X_i} \cdot \frac{1}{\tilde{\lambda}^2} \cdot \left((X_i - \tilde{\lambda})^2 - X_i \right) \right|^3 \\
&= \left| e^{\lambda_0 - \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{X_i} \cdot \frac{1}{\tilde{\lambda}^2} \cdot \left(X_i^2 - X_i(2\tilde{\lambda} + 1) + \tilde{\lambda}^2 \right) \right|^3 \\
&\leq \left| e^{\lambda_0 - c} \left(\frac{d}{\lambda_0} \right)^{X_i} \cdot \frac{1}{c^2} \cdot \left(X_i^2 + X_i(2d + 1) + d^2 \right) \right|^3 \quad \text{da } \tilde{\lambda} \in (\lambda_0, \lambda) \subset [c, d] \\
&= \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot \frac{1}{c^6} \cdot \left(X_i^2 + X_i(2d + 1) + d^2 \right)^3 \right| \\
&\leq \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot a_2 e^{X_i} \right| \quad \text{mit } a_2 \in \mathbb{R}_+ \\
&= \left| aa_2 e^{X_i \left(3 \log\left(\frac{d}{\lambda_0}\right) + 1 \right)} \right|
\end{aligned}$$

$$\begin{aligned}
|Z_i''(\lambda)|^2 &= \left| \frac{f^{(4)}(X_i; \tilde{\lambda})}{f(X_i; \lambda_0)} \right|^2 \quad \text{mit } \tilde{\lambda} \in (\lambda_0, \lambda) \text{ für o.B.d.A. } \lambda_0 \leq \lambda \\
&= \left| \frac{e^{-\tilde{\lambda}} \tilde{\lambda}^{X_i-4} ((X_i - \tilde{\lambda})^4 - X_i(6X_i^2 - (12\tilde{\lambda} + 11)X_i + (6\tilde{\lambda}^2 + 8\tilde{\lambda} + 6)))}{e^{-\lambda_0} \lambda_0^{X_i}} \right|^2 \\
&= \left| e^{\lambda_0 - \tilde{\lambda}} \left(\frac{\tilde{\lambda}}{\lambda_0} \right)^{X_i} \cdot \frac{1}{\tilde{\lambda}^4} \cdot ((X_i - \tilde{\lambda})^4 - X_i(6X_i^2 - (12\tilde{\lambda} + 11)X_i + \right. \\
&\quad \left. + (6\tilde{\lambda}^2 + 8\tilde{\lambda} + 6))) \right|^2 \\
&\leq \left| e^{\lambda_0 - c} \left(\frac{d}{\lambda_0} \right)^{X_i} \cdot \frac{1}{c^4} \cdot (X_i^4 + (4d + 6)X_i^3 + (6d^2 + 12d + 11)X_i^2 + \right. \\
&\quad \left. + (4d^3 + 6d^2 + 8d + 6)X_i + d^4) \right|^2 \\
&= \left| \tilde{a} \cdot e^{X_i \cdot 2 \log\left(\frac{d}{\lambda_0}\right)} \cdot \frac{1}{c^8} \cdot (X_i^4 + (4d + 6)X_i^3 + (6d^2 + 12d + 11)X_i^2 + \right. \\
&\quad \left. + (4d^3 + 6d^2 + 8d + 6)X_i + d^4)^2 \right| \quad \text{mit } \tilde{a} = e^{2(\lambda_0 - c)} \\
&\leq \left| \tilde{a} \cdot e^{X_i \cdot 2 \log\left(\frac{d}{\lambda_0}\right)} \cdot a_3 e^{X_i} \right| \quad \text{mit } a_3 \in \mathbb{R}_+ \\
&\leq \left| a \cdot e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \cdot a_3 e^{X_i} \right| \quad \text{da } a \geq \tilde{a} \text{ und } e^{X_i \cdot 3 \log\left(\frac{d}{\lambda_0}\right)} \geq e^{X_i \cdot 2 \log\left(\frac{d}{\lambda_0}\right)} \\
&= \left| a a_3 e^{X_i (3 \log\left(\frac{d}{\lambda_0}\right) + 1)} \right|
\end{aligned}$$

Die Funktion

$$g(X_i) = \left| a \max\{a_1, a_2, a_3\} e^{X_i (3 \log\left(\frac{d}{\lambda_0}\right) + 1)} \right|$$

majorisiert $|Y_i(\lambda)|^3$, $|Z_i(\lambda)|^3$ und $|Z_i''(\lambda)|^2$. Wegen $E \exp(tX_i) = e^{\lambda(e^t - 1)}$, $t \in \mathbb{R}$, ist g auch integrierbar.

Zusammen mit den hier nicht explizit nachgerechneten Bedingungen folgt, dass Poissonverteilungen die Regularitätsbedingungen erfüllen.

In Bezug auf Normalverteilungen wird im Folgenden zwischen Lokations- und Skalenfamilien unterschieden. Bei Lokationsfamilien ist die Varianz σ_0^2 fix (o.B.d.A. wird $\sigma_0^2 = 1$ gewählt) und der Erwartungswert μ ist variabel. Bei Skalenfamilien dagegen ist die Varianz variabel und der Erwartungswert wird o.B.d.A. mit $\mu_0 = 0$ fest gewählt.

Beispiel 2.1.3 Zunächst werden normalverteilte Zufallsvariablen X_i mit fester Varianz und variablem Erwartungswert betrachtet. Für $\sigma_0^2 = 1$ besitzen sie die Dichte

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}.$$

Die Zufallsvariablen 2.2 sind dann von der Form

$$\begin{aligned} Y_i(\mu) &= \frac{e^{X_i(\mu-\mu_0)-\frac{1}{2}(\mu^2-\mu_0^2)} - 1}{\mu - \mu_0} \\ Y_i(\mu_0) &= X_i - \mu_0 \\ Z_i(\mu) &= \frac{e^{X_i(\mu-\mu_0)-\frac{1}{2}(\mu^2-\mu_0^2)} - 1}{(\mu - \mu_0)^2} - \frac{X_i - \mu_0}{\mu - \mu_0} \\ Z_i(\mu_0) &= \frac{1}{2} ((X_i - \mu_0)^2 - 1). \end{aligned}$$

A1. Der Parameterraum $\Theta = [c, d]$ mit $c, d \in \mathbb{R}, c \leq \mu_1, \mu_2 \leq d$ ist kompakt.

A2. (a) Es gilt

$$\begin{aligned} & E \left| \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2} \right) \right| \\ &= E \left| \log \frac{1}{\sqrt{2\pi}} - \frac{1}{2}(X - \mu)^2 \right| \\ &\leq E \left| \log \frac{1}{\sqrt{2\pi}} \right| + \frac{1}{2} E(X - \mu)^2 \\ &= \left| \log \frac{1}{\sqrt{2\pi}} \right| + \frac{1}{2} < \infty, \end{aligned}$$

da $E(X - \mu)^2 = \sigma_0^2 = 1$ ist.

A3. Der Träger der Normalverteilung ist \mathbb{R} und somit von μ unabhängig. Die Dichte $f(x; \mu)$ ist zweimal stetig differenzierbar bezüglich μ :

$$f'(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} (x - \mu)$$

$$f''(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} ((x - \mu)^2 - 1).$$

A4. (a) Ein Beweis zur Identifizierbarkeit von endlichen Mischungen von Normalverteilungen findet sich bei Teicher (1963). Hier wird erneut Satz C.1 verwendet.

Sei $f(x)$ die Dichte einer standardnormalverteilten Zufallsvariable und $\mathcal{F} = \{F_\mu : F_\mu = F(x, \mu) = \int_{-\infty}^x f(t - \mu) dt\}$ eine Familie von Normalverteilungen mit variablem Erwartungswert μ und fester Varianz $\sigma_0^2 = 1$. \mathcal{F} lässt sich durch

$$F_{\mu_1} = F(x, \mu_1) \prec F(x, \mu_2) = F_{\mu_2},$$

falls $\mu_1 < \mu_2$, ordnen. Die zweiseitige Laplacetransformation der Normalverteilung ist für $\sigma_0^2 = 1$ gegeben durch

$$\phi(t; \mu, 1) = e^{\frac{t^2}{2} - \mu t}.$$

Sie ist bijektiv und linear und für beliebige Erwartungswerte μ_1 und μ_2 besitzen die entsprechenden Funktionen ϕ_1 und ϕ_2 denselben Definitionsbereich $S_\phi = (-\infty, \infty)$. Für $t_1 = +\infty$ gilt

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t \rightarrow +\infty} \frac{e^{\frac{t^2}{2} - \mu_2 t}}{e^{\frac{t^2}{2} - \mu_1 t}} = \lim_{t \rightarrow +\infty} \frac{e^{-\mu_2 t}}{e^{-\mu_1 t}} = \lim_{t \rightarrow +\infty} e^{(\mu_1 - \mu_2)t} = 0,$$

da $\mu_1 < \mu_2$. Somit sind nach Teicher (1963) alle endlichen Mischungen von \mathcal{F} identifizierbar.

(b) Um zu zeigen, dass endliche Mischungen aus der Lokationsfamilie von Normalverteilungen $\mathcal{F} = \{F_\mu\}$ stark identifizierbar sind, wird Satz C.2 von Chen (1995) verwendet. Es gilt

$$\lim_{x \rightarrow \pm\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} = 0$$

und

$$\lim_{x \rightarrow \pm\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} (x - \mu) = 0,$$

da $e^{-\frac{1}{2}(x-\mu)^2}$ schneller gegen 0 geht, als $(x - \mu)$ wächst. Somit folgt nach Satz C.2 die starke Identifizierbarkeit für endliche Mischungen aus \mathcal{F} .

A5.

$$\begin{aligned} |Y_i(\mu)|^3 &= \left| \frac{f'(X_i; \tilde{\mu})}{f(X_i; \mu_0)} \right|^3 \quad \text{mit } \tilde{\mu} \in (\mu_0, \mu) \text{ für o.B.d.A } \mu_0 \leq \mu \\ &= \left| \frac{e^{-\frac{1}{2}(X_i - \tilde{\mu})^2} (X_i - \tilde{\mu})}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{e^{-\frac{1}{2}(X_i - \mu_0)^2}} \right|^3 \\ &= \left| e^{\frac{1}{2}((X_i - \mu_0)^2 - (X_i - \tilde{\mu})^2)} (X_i - \tilde{\mu}) \right|^3 \\ &= \left| e^{X_i(\tilde{\mu} - \mu_0)} \cdot e^{\frac{1}{2}(\mu_0^2 - \tilde{\mu}^2)} (X_i - \tilde{\mu}) \right|^3 \\ &= \left| e^{\frac{3}{2}(\mu_0^2 - \tilde{\mu}^2)} (X_i e^{X_i(\tilde{\mu} - \mu_0)} - \tilde{\mu} e^{X_i(\tilde{\mu} - \mu_0)}) \right|^3 \\ &\leq \left| e^{\frac{3}{2}(\mu_0^2 - c^2)} (X_i e^{X_i(d - \mu_0)} - c e^{X_i(d - \mu_0)}) \right|^3 \quad \text{da } \tilde{\mu} \in (\mu_0, \mu) \subset [c, d] \\ &= \left| a (X_i e^{X_i(d - \mu_0)} - c e^{X_i(d - \mu_0)}) \right|^3 \quad \text{mit } a := e^{\frac{3}{2}(\mu_0^2 - c^2)} \\ &\leq \left| a (e^{X_i} e^{X_i(d - \mu_0)} - c e^{X_i(d - \mu_0)}) \right|^3 \\ &= \left| a e^{X_i \cdot 3(d - \mu_0)} (e^{X_i} - c)^3 \right| \end{aligned}$$

$$\begin{aligned}
|Z_i(\mu)|^3 &= |Y_i'(\tilde{\mu})|^3 \quad \text{mit } \tilde{\mu} \in (\mu_0, \mu) \text{ für o.B.d.A. } \mu_0 \leq \mu \\
&= \left| \frac{f''(X_i; \tilde{\mu})}{f(X_i; \mu_0)} \right|^3 \\
&= \left| \frac{e^{-\frac{1}{2}(X_i - \tilde{\mu})^2} ((X_i - \tilde{\mu})^2 - 1)}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{e^{-\frac{1}{2}(X_i - \mu_0)^2}} \right|^3 \\
&= \left| e^{\frac{1}{2}((X_i - \mu_0)^2 - (X_i - \tilde{\mu})^2)} ((X_i - \tilde{\mu})^2 - 1) \right|^3 \\
&\leq \left| a (e^{X_i(d - \mu_0)} \cdot (X_i^2 + 2X_id + d^2) - e^{X_i(d - \mu_0)})^3 \right| \quad \text{da } \tilde{\mu} \in (\mu_0, \mu) \subset [c, d] \\
&\leq \left| a (e^{X_i(d - \mu_0)} \cdot (X_i^2 + 2X_id + d^2 + 1))^3 \right| \\
&= \left| a (e^{X_i(d - \mu_0)} \cdot a_1 e^{X_i})^3 \right| \quad \text{mit } a_1 := 2! + 2d + d^2 + 1 \\
&= \left| \tilde{a} e^{X_i(3((d - \mu_0) + 1))} \right| \quad \text{mit } \tilde{a} := a a_1^3 \geq a
\end{aligned}$$

$$\begin{aligned}
|Z_i''(\mu)|^2 &= \left| \frac{f^{(4)}(X_i; \tilde{\mu})}{f(X_i; \mu_0)} \right|^2 \quad \text{mit } \tilde{\mu} \in (\mu, \mu_0) \text{ für o.B.d.A. } \mu \leq \mu_0 \\
&= \left| \frac{e^{-\frac{1}{2}(X_i - \tilde{\mu})^2} ((X_i - \tilde{\mu})^4 - 6(X_i - \tilde{\mu})^2 + 3)}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{e^{-\frac{1}{2}(X_i - \mu_0)^2}} \right|^2 \\
&= \left| e^{\frac{1}{2}((X_i - \mu_0)^2 - (X_i - \tilde{\mu})^2)} \cdot ((X_i - \tilde{\mu})^4 - 6(X_i - \tilde{\mu})^2 + 3) \right|^2 \\
&= \left| e^{\mu_0^2 - \tilde{\mu}^2} \cdot e^{X_i \cdot 2(\tilde{\mu} - \mu_0)} \cdot ((X_i - \tilde{\mu})^4 - 6(X_i - \tilde{\mu})^2 + 3)^2 \right| \\
&\leq \left| \bar{a} e^{X_i \cdot 2(d - \mu_0)} ((X_i - \tilde{\mu})^4 - 6(X_i - \tilde{\mu})^2 + 3)^2 \right| \quad \text{da } \tilde{\mu} \in (\mu_0, \mu) \subset [c, d] \\
&\quad \text{und mit } \bar{a} := e^{\mu_0^2 - c^2} \\
&\leq \left| \bar{a} e^{X_i \cdot 2(d - \mu_0)} \cdot a_2 e^{2X_i} \right| \quad \text{mit geeigneter Konstante } a_2 \text{ (analog zu } a_1) \\
&= \left| \bar{a} a_2 e^{X_i \cdot 2((d - \mu_0) + 1)} \right|
\end{aligned}$$

Falls $|d - \mu_0| \geq 1$ gilt, so ist eine majorisierende Funktion für $|Y_i(\mu)|^3$, $|Z_i(\mu)|^3$ und $|Z_i''(\mu)|^2$ gegeben durch

$$g(X_i) = \left| C e^{X_i(3((d - \mu_0) + 1))} \right|$$

mit geeigneter Konstante C . Wegen $E \exp(tX_i) = e^{\mu_0 t + \frac{t^2}{2}}$, $t \in \mathbb{R}$, ist $g(X_i)$ integrierbar.

Damit erfüllen auch Normalverteilungen mit fixer Varianz und variablem Erwartungswert die Regularitätsbedingungen.

Beispiel 2.1.4 Für Normalverteilungen mit festem Erwartungswert $\mu_0 = 0$ und variabler Varianz σ^2 ist die Dichte von der Form

$$f(x; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}.$$

Die Zufallsvariablen 2.2 sind gegeben durch

$$\begin{aligned} Y_i(\sigma) &= \frac{\sigma_0 e^{-\frac{1}{2}X_i^2\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}}{\sigma(\sigma - \sigma_0)} - \frac{1}{\sigma - \sigma_0} \\ Y_i(\sigma_0) &= \frac{(X_i^2 - \sigma_0^2)}{\sigma_0^3} \\ Z_i(\sigma) &= \frac{\sigma_0 e^{-\frac{1}{2}X_i^2\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}}{\sigma(\sigma - \sigma_0)^2} - \frac{1}{(\sigma - \sigma_0)^2} - \frac{(X_i^2 - \sigma_0^2)}{\sigma_0^3(\sigma - \sigma_0)} \\ Z_i(\sigma_0) &= \frac{\frac{1}{2}(X_i^4 - 5\sigma_0^2 X_i^2 + 2\sigma_0^4)}{\sigma_0^6}. \end{aligned}$$

A1. Der Parameterraum $\Theta = [c, d]$ mit $c, d \in \mathbb{R}$, $0 < c < d$ ist kompakt.

A2. (a) Es gilt

$$\begin{aligned} & E \left| \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{X}{\sigma}\right)^2} \right) \right| \\ &= E \left| \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2} \left(\frac{X}{\sigma} \right)^2 \right| \\ &\leq E \left| \log \frac{1}{\sqrt{2\pi\sigma^2}} \right| + \frac{1}{2} E \left(\frac{X}{\sigma} \right)^2 \\ &= E \left| -\log \sqrt{2\pi\sigma^2} \right| + \frac{1}{2} \frac{EX^2}{\sigma^2} \end{aligned}$$

$$\begin{aligned}
&\leq E \left| \log \sqrt{2\pi} \right| + E |\log \sigma| + \frac{1}{2} \frac{\text{Var} X - (EX)^2}{\sigma^2} \\
&= |\log \sqrt{2\pi}| + |\log \sigma| + \frac{1}{2} \\
&\leq |\log \sqrt{2\pi}| + \max\{|\log c|, |\log d|\} + \frac{1}{2} < \infty.
\end{aligned}$$

A3. Der Träger \mathbb{R} der Normalverteilung mit festem Erwartungswert und variabler Varianz ist unabhängig von σ . Die Dichte $f(x; \sigma)$ ist zweimal stetig differenzierbar bezüglich σ :

$$\begin{aligned}
f'(x; \sigma) &= \frac{1}{\sqrt{2\pi}\sigma^4} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} (x^2 - \sigma^2) \\
f''(x; \sigma) &= \frac{1}{\sqrt{2\pi}\sigma^7} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} (x^4 - 5\sigma^2 x^2 + 2\sigma^4).
\end{aligned}$$

A4. (a) Einen Beweis zur Identifizierbarkeit liefert Teicher (1963). Sei $f(x)$ die Dichte einer standardnormalverteilten Zufallsvariable X und $\mathcal{F} = \{F_\sigma : F_\sigma = F(x, \sigma) = \frac{1}{\sigma} \int_{-\infty}^x f\left(\frac{t-\mu_0}{\sigma}\right) dt\}$ eine Familie von Normalverteilungen mit festem Erwartungswert $\mu_0 = 0$ und variabler Varianz σ^2 . Um die Identifizierbarkeit von endlichen Mischungen von \mathcal{F} zu zeigen, wird erneut Satz C.1 von Teicher (1963) verwendet. \mathcal{F} lässt sich durch

$$F_{\sigma_1} = F(x, \sigma_1) \prec F(x, \sigma_2) = F_{\sigma_2},$$

falls $\sigma_1 > \sigma_2$, ordnen. Die zweiseitige Laplacetransformation

$$\phi(t; 0, \sigma^2) = e^{-\frac{\sigma^2 t^2}{2}}$$

ist linear und bijektiv. Für beliebige Varianzen σ_1 und σ_2 besitzen die entsprechenden Funktionen ϕ_1 und ϕ_2 denselben Definitionsbereich $S_\phi = (-\infty, \infty)$. Für $t_1 = +\infty$ gilt

$$\lim_{t \rightarrow +\infty} \frac{\phi_2(t)}{\phi_1(t)} = \frac{e^{-\frac{\sigma_2^2 t^2}{2}}}{e^{-\frac{\sigma_1^2 t^2}{2}}} = \lim_{t \rightarrow +\infty} e^{(\sigma_2^2 - \sigma_1^2) \frac{t^2}{2}} = 0,$$

da $\sigma_1 > \sigma_2$ ist. Aus Satz C.1 folgt somit, dass alle endlichen Mischungen aus \mathcal{F} identifizierbar sind.

(b) Die starke Identifizierbarkeit von endlichen Mischungen aus der Familie $\mathcal{F} = \{F_\sigma\}$ lässt sich mit Satz C.3 von Chen (1995) zeigen. Es gilt

$$\lim_{x \rightarrow \pm\infty} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = 0$$

und

$$\lim_{x \rightarrow \pm\infty} \frac{1}{\sqrt{2\pi}\sigma^3} x e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = 0,$$

da $e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ schneller gegen 0 geht als x gegen ∞ . Also sind nach Satz C.3 endliche Mischungen aus \mathcal{F} stark identifizierbar.

A5.

$$\begin{aligned} |Y_i(\sigma)|^3 &= \left| \frac{f'(X_i; \tilde{\sigma})}{f(X_i; \sigma_0)} \right|^3 \quad \text{mit } \tilde{\sigma} \in (\sigma_0, \sigma) \text{ für o.B.d.A. } \sigma_0 \leq \sigma \\ &= \left| \frac{e^{-\frac{1}{2}\left(\frac{X_i}{\tilde{\sigma}}\right)^2} (X_i^2 - \tilde{\sigma}^2)}{\sqrt{2\pi}\tilde{\sigma}^4} \cdot \frac{\sqrt{2\pi}\sigma_0}{e^{-\frac{1}{2}\left(\frac{X_i}{\sigma_0}\right)^2}} \right|^3 \\ &= \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{\tilde{\sigma}}\right)^2\right)} \cdot (X_i^2 - \tilde{\sigma}^2) \cdot \frac{\sigma_0}{\tilde{\sigma}^4} \right|^3 \\ &\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot (X_i^2 - c^2) \cdot \frac{\sigma_0}{c^4} \right|^3 \quad \text{da } \tilde{\sigma} \in (\sigma_0, \sigma) \subset [c, d] \\ &\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot \max\{1, c^2\} \cdot (X_i^2 + 1) \cdot \frac{\sigma_0}{c^4} \right|^3 \\ &= \left| e^{\frac{3}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot \max\{1, c^2\}^3 \cdot (X_i^2 + 1)^3 \cdot \left(\frac{\sigma_0}{c^4}\right)^3 \right|^3 \\ &= \left| a_1 e^{\frac{3}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot (X_i^2 + 1)^3 \right|^3 \quad \text{mit } a_1 := \max\{1, c^2\}^3 \cdot \left(\frac{\sigma_0}{c^4}\right)^3 \end{aligned}$$

$$\begin{aligned}
|Z_i(\sigma)|^3 &= |Y'_i(\tilde{\sigma})|^3 \quad \text{mit } \tilde{\sigma} \in (\sigma_0, \sigma) \text{ für o.B.d.A. } \sigma_0 \leq \sigma \\
&= \left| \frac{f''(X_i; \tilde{\sigma})}{f(X_i; \sigma_0)} \right|^3 \\
&= \left| \frac{e^{-\frac{1}{2}\left(\frac{X_i}{\tilde{\sigma}}\right)^2} (X_i^4 - 5\tilde{\sigma}^2 X_i^2 + 2\tilde{\sigma}^4) \cdot \frac{\sqrt{2\pi}\sigma_0}{e^{-\frac{1}{2}\left(\frac{X_i}{\sigma_0}\right)^2}}}{\sqrt{2\pi}\tilde{\sigma}^7} \right|^3 \\
&= \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{\tilde{\sigma}}\right)^2\right)} \cdot (X_i^4 - 5\tilde{\sigma}^2 X_i^2 + 2\tilde{\sigma}^4) \cdot \frac{\sigma_0}{\tilde{\sigma}^7} \right|^3 \\
&\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot (X_i^4 + 5d^2 X_i^2 + 2d^4) \cdot \frac{\sigma_0}{c^7} \right|^3 \quad \text{da } \tilde{\sigma} \in (\sigma_0, \sigma) \subset [c, d] \\
&\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot \max\{1, 5d^2, 2d^4\} \cdot (X_i^4 + X_i^2 + 1) \cdot \frac{\sigma_0}{c^7} \right|^3 \\
&= \left| e^{\frac{3}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot \max\{1, 5d^2, 2d^4\}^3 \cdot (X_i^4 + X_i^2 + 1)^3 \cdot \left(\frac{\sigma_0}{c^7}\right)^3 \right| \\
&= \left| a_2 e^{\frac{3}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot (X_i^4 + X_i^2 + 1)^3 \right| \quad \text{mit } a_2 := \max\{1, 5d^2, 2d^4\}^3 \cdot \left(\frac{\sigma_0}{c^7}\right)^3
\end{aligned}$$

$$\begin{aligned}
|Z_i''(\sigma)|^2 &= \left| \frac{f^{(4)}(X_i; \tilde{\sigma})}{f(X_i; \sigma_0)} \right|^2 \quad \text{mit } \tilde{\sigma} \in (\sigma_0, \sigma) \text{ für o.B.d.A. } \sigma_0 \leq \sigma \\
&= \left| \frac{e^{-\frac{1}{2}\left(\frac{X_i}{\tilde{\sigma}}\right)^2} (X_i^8 - 22\tilde{\sigma}^2 X_i^6 + 123\tilde{\sigma}^4 X_i^4 - 168\tilde{\sigma}^6 X_i^2 + 24\tilde{\sigma}^8)}{\sqrt{2\pi}\tilde{\sigma}^{13}} \cdot \frac{\sqrt{2\pi}\sigma_0}{e^{-\frac{1}{2}\left(\frac{X_i}{\sigma_0}\right)^2}} \right|^2 \\
&= \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{\tilde{\sigma}}\right)^2\right)} \cdot (X_i^8 - 22\tilde{\sigma}^2 X_i^6 + 123\tilde{\sigma}^4 X_i^4 - 168\tilde{\sigma}^6 X_i^2 + 24\tilde{\sigma}^8) \cdot \frac{\sigma_0}{\tilde{\sigma}^{13}} \right|^2 \\
&\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot (X_i^8 + 22d^2 X_i^6 + 123d^4 X_i^4 + 168d^6 X_i^2 + 24d^8) \cdot \frac{\sigma_0}{c^{13}} \right|^2 \\
&\leq \left| e^{\frac{1}{2}\left(\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2\right)} \cdot \max\{1, 22d^2, 123d^4, 168d^6, 24d^8\} \right. \\
&\quad \left. \cdot (X_i^8 + X_i^6 + X_i^4 + X_i^2 + 1) \cdot \frac{\sigma_0}{c^{13}} \right|^2 \\
&= \left| a_3 e^{\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2} \cdot (X_i^8 + X_i^6 + X_i^4 + X_i^2 + 1)^2 \right| \\
&\quad \text{mit } a_3 := \max\{1, 22d^2, 123d^4, 168d^6, 24d^8\}^2 \cdot \left(\frac{\sigma_0}{c^{13}}\right)^2
\end{aligned}$$

Für $d < \sqrt{\frac{3}{2}}\sigma_0$ existieren die Erwartungswerte

$$E \left(X_i^k e^{\frac{3}{2}X_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{d^2}\right)} \right), \quad k \in \mathbb{N}$$

und

$$E \left(X_i^k e^{X_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{d^2}\right)} \right), \quad k \in \mathbb{N}.$$

Eine integrierbare Funktion, die $|Y_i(\sigma)|^3$, $|Z_i(\sigma)|^3$ und $|Z_i''(\sigma)|^2$ majorisiert, ist dann gegeben durch

$$g(X_i) = \left| e^{\left(\frac{X_i}{\sigma_0}\right)^2 - \left(\frac{X_i}{d}\right)^2} \cdot \left(\sum_{k=0}^8 X_i^{2k} \right) \cdot \max\{a_1, a_2, a_3\} \right|.$$

Somit erfüllen auch Normalverteilungen aus Skalenfamilien die Regularitätsbedingungen.

Beispiel 2.1.5 Eine binomialverteilte Zufallsvariable besitzt die Dichte

$$f(x; p) = \binom{n}{x} p^x (1-p)^{n-x},$$

wobei die Anzahl n der Bernoulliexperimente fix sei. Die Zufallsvariablen 2.2 sind dann

$$\begin{aligned} Y_i(p) &= \frac{1}{p-p_0} \left(\left(\frac{p}{p_0} \right)^{X_i} \left(\frac{1-p}{1-p_0} \right)^{n-X_i} - 1 \right) \\ Y_i(p_0) &= \frac{X_i - p_0 n}{p_0(1-p_0)} \\ Z_i(p) &= \frac{1}{p_0(p-p_0)^2} \left(\frac{p^{X_i}}{p_0^{X_i-1}} \left(\frac{1-p}{1-p_0} \right)^{n-X_i} - p_0 - \frac{(X_i - p_0 n)(p-p_0)}{1-p_0} \right) \\ Z_i(p_0) &= \frac{X_i^2 - X_i(2p_0 n - 2p_0 + 1) + p_0^2 n(n-1)}{2p_0^2(1-p_0)^2}. \end{aligned}$$

A1. Der Parameterraum $\Theta = [0, 1]$ ist kompakt.

A2. Es gilt

$$\begin{aligned} & E \left| \log \left(\binom{n}{X} p^X (1-p)^{n-X} \right) \right| \\ &= E \left| \log \frac{n!}{X!(n-X)!} + X \log p + (n-X) \log(1-p) \right| \\ &\leq E |\log n!| + E |\log X!| + E |\log(n-X)!| + E |X \log p| + E |(n-X) \log(1-p)| \\ &= \log n! + E |\log X!| + E |\log(n-X)!| + |\log p| n p + E |(n-X) \log(1-p)|. \end{aligned}$$

Da $p \in [0, 1]$ und n endlich ist, sind $\log n!$ und $|\log p| n p$ endlich. Mit der Jensenschen Ungleichung lässt sich $E |\log X!|$ durch $\log EX!$ nach oben abschätzen. Es ist

$$EX! = \sum_{k=0}^n k! \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=0}^n \frac{n!}{(n-k)!} p^k (1-p)^{n-k} < \infty$$

und somit ist auch $\log EX!$ endlich.

Analog lässt sich zeigen, dass $E|\log(n - X)!| < \infty$ ist.

Es gilt weiterhin

$$\begin{aligned} & E|(n - X) \log(1 - p)| \\ & \leq |\log(1 - p)|(E|n| + E|X|) \\ & = |\log(1 - p)|(n + np) < \infty. \end{aligned}$$

Folglich ist $E|\log \binom{n}{X} p^X (1 - p)^{n-X}| < \infty$.

A3. Der Träger $\{0, \dots, n\}$ der Binomialverteilung ist unabhängig von $p \in (0, 1)$ und $f(x; p)$ ist zweimal stetig differenzierbar bezüglich p :

$$\begin{aligned} f'(x; p) &= \binom{n}{x} p^{x-1} (1 - p)^{n-x-1} (x - pn) \\ f''(x; p) &= \binom{n}{x} p^{x-2} (1 - p)^{n-x-2} (x^2 - x(2pn - 2p + 1) + p^2 n(n - 1)). \end{aligned}$$

A4. (a) Den Beweis zur Identifizierbarkeit von endlichen Mischungen von Binomialverteilungen liefern Titterton et al. (1985). Dabei verwenden sie Satz C.4 von Teicher (1963), welcher besagt, dass Mischungen mit maximal k Komponenten der Familie

$$\mathcal{F} = \{F(x; n, p), 0 < p < 1, n \in \mathbb{N} \text{ fest}\}$$

von Binomialverteilungen genau dann identifizierbar sind, wenn $n \geq 2k - 1$ ist.

Seien

$$F(x; n, p) = \sum_{j=1}^{c_1} \pi_j F_j(x) = \sum_{j=1}^{c_2} \tilde{\pi}_j \tilde{F}_j(x), \quad x = 1, \dots, n$$

mit

$$\begin{aligned} 0 &< \pi_j < 1, & j &= 1, \dots, c_1, \\ 0 &< \tilde{\pi}_j < 1, & j &= 1, \dots, c_2, \\ \sum_j \pi_j &= \sum_j \tilde{\pi}_j = 1 \text{ und } c_1, c_2 \leq k \end{aligned}$$

zwei verschiedene Darstellungen einer Mischung von Binomialverteilungen. Dann ist

$$\sum_{j=1}^{k'} \lambda_j F_j^*(x) = 0, \quad x = 1, \dots, n. \quad (2.4)$$

Dabei seien $\{F_j^*\}$ die k' verschiedenen Elemente aus der Menge $\{F_1, \dots, F_{c_1}, \tilde{F}_1, \dots, \tilde{F}_{c_2}\}$.

Die wahrscheinlichkeitserzeugende Funktion der Binomialverteilung ist gegeben durch $g(z) = (1 - p + pz)^n$. Im Folgenden wird sie jedoch an der Stelle $z+1$ betrachtet. Es ist dann $g(z+1) = (1+pz)^n$ und 2.4 ist äquivalent zu

$$\sum_{j=1}^{k'} \lambda_j (1 + p_j^* z)^n = 0 \quad \text{für alle } z \in \mathbb{R}.$$

Nach dem binomischen Lehrsatz ist

$$\begin{aligned} & \sum_{j=1}^{k'} \lambda_j (1 + p_j^* z)^n = 0 \quad \text{für alle } z \in \mathbb{R} \\ \Leftrightarrow & \sum_{j=1}^{k'} \lambda_j \sum_{i=0}^n \binom{n}{i} (p_j^* z)^i = 0 \quad \text{für alle } z \in \mathbb{R} \\ \Leftrightarrow & \sum_{i=0}^n \sum_{j=1}^{k'} \lambda_j \binom{n}{i} p_j^{*i} z^i = 0 \quad \text{für alle } z \in \mathbb{R}. \end{aligned}$$

Man erhält auf der linken Seite der Gleichung ein Polynom maximal n -ten Grades mit Koeffizienten $\sum_{j=1}^{k'} \lambda_j \binom{n}{i} p_j^{*i}$, $i = 1, \dots, n$. Die Null auf der rechten Seite lässt sich als Nullpolynom gleichen Grades auffassen. Die beiden Polynome stimmen an mindestens $n+1$ Stellen überein, sodass nach dem Identitätssatz für Polynome folgt, dass sie identisch sind. Insbesondere gilt für die Koeffizienten

$$\sum_{j=1}^{k'} \lambda_j \binom{n}{i} p_j^{*i} = 0 \quad \text{für } i = 0, \dots, n.$$

Hieraus folgt, da der Binomialkoeffizient echt größer 0 ist, dass

$$\sum_{j=1}^{k'} \lambda_j p_j^{*i} = 0 \quad \text{für } i = 0, \dots, n.$$

Nach Yakowitz und Spragins (1968) sind endliche Mischungen von Verteilungsfunktionen genau dann identifizierbar, wenn die Komponenten linear unabhängig sind (Satz C.5). Für $n < k' - 1$ ist Identifizierbarkeit ausgeschlossen, da eine von 0 verschiedene Lösung für $(\lambda_1, \dots, \lambda_{k'})$ existiert. Für $n \geq k' - 1$ ist aber $\lambda_1 = \dots = \lambda_{k'} = 0$ die einzige Lösung. Da der Beweis für beliebige Paare (c_1, c_2) gilt, sind alle endlichen Mischungen aus maximal k Komponenten aus \mathcal{F} für $n \geq k - 1$ identifizierbar.

Somit erfüllen auch Binomialverteilungen die Regularitätsbedingungen.

Mithilfe der asymptotischen Verteilung einer Teststatistik lassen sich die kritischen Werte (Quantile) für den angewandten Test ermitteln. Dies erweist sich jedoch als problematisch für eine Verteilung wie in Satz 2.1.1. Hierbei ist einerseits die Berechnung des Supremums eines Gaußprozesses nicht sehr einfach und andererseits sind durch die Abhängigkeit der Verteilung von θ und der Kernfunktion $f(x; \theta)$ für verschiedene Parameter und Kernfunktionen stets neue Berechnungen erforderlich. Diese Probleme stellen sich nicht, wenn man die Teststatistik R_n so anpasst, dass sie in Verteilung gegen eine Zufallsvariable konvergiert, deren Verteilung einfacher ist und nicht von θ abhängt. Dies führt zu dem von Chen et al. (2001) vorgeschlagenen modifizierten Likelihood-Quotienten-Test.

2.1.2 Der modifizierte Likelihood-Quotienten-Test

Die Abhängigkeit der asymptotischen Nullverteilung der Likelihood-Quotienten-Test-Statistik von θ und der Kernfunktion erschwert die Berechnung der kritischen Werte. Eine dafür geeignetere Verteilung ergibt sich bei der Anwendung des modifizierten Likelihood-Quotienten-Tests (MLQT) von Chen et al. (2001).

Wie bereits in Kapitel 2.1.1 erwähnt, sind zwei der von Ferguson (1996) angegebenen Regularitätsbedingungen im betrachteten Mischungsmodell 2.1

nicht erfüllt: Die Hypothese liegt für $\alpha = 0$ oder $\alpha = 1$ auf dem Rand des Parameterraums und die Mischungsdichte ist unter der Hypothese nicht identifizierbar. Mit dem modifizierten Likelihood-Quotienten-Test lässt sich jedoch das Problem der Nichtidentifizierbarkeit der Parameter durch das Addieren einer Penalty-Funktion zur Log-Likelihood-Funktion umgehen. Anstelle von $l_n(\alpha, \theta_1, \theta_2)$ wird im MLQT die Funktion

$$pl_n(\alpha, \theta_1, \theta_2) = l_n(\alpha, \theta_1, \theta_2) + p(\alpha)$$

betrachtet, wobei $p : [0, 1] \rightarrow \mathbb{R}_{\leq 0}$ eine stetige Funktion ist, die für $\alpha \rightarrow 0$ oder 1 gegen $-\infty$ strebt und für $\alpha = 0.5$ ihr Maximum annimmt. Seien

$$(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2) = \operatorname{argmax}_{[0,1] \times \Theta^2} pl_n(\alpha, \theta_1, \theta_2) \quad \text{und} \quad \tilde{\theta}_0 = \operatorname{argmax}_{\Theta} pl_n(1/2, \theta, \theta)$$

die Maximum-Likelihood-Schätzer unter der Alternative und der Hypothese. Die Teststatistik des MLQT lautet

$$M_n = 2\{pl_n(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2) - pl_n(1/2, \tilde{\theta}_0, \tilde{\theta}_0)\}.$$

Durch die Penalty-Funktion $p(\alpha)$ wird verhindert, dass es einen Maximum-Likelihood-Schätzer für $\tilde{\alpha} = 0$ oder $\tilde{\alpha} = 1$ gibt.

Satz 2.1.6 (Chen et al., 2001) *Gelten die Regularitätsbedingungen A1-A5 und ist die Hypothese erfüllt, so gilt*

$$M_n \xrightarrow{d} \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2.$$

Dabei ist χ_1^2 eine Chi-Quadrat-Verteilung mit einem Freiheitsgrad und χ_0^2 eine Einpunktverteilung in 0.

Es gibt verschiedene Möglichkeiten die Penalty-Funktion zu wählen. Chen et al. (2001) schlagen die Funktion

$$p(\alpha) = C \log\{4\alpha(1 - \alpha)\}$$

mit positiver Konstante C vor. Im Folgenden soll jedoch die von Li et al. (2008) entwickelte und effektivere Penalty-Funktion

$$p(\alpha) = C^* \log(1 - |1 - 2\alpha|)$$

mit $C^* = 1$ verwendet werden.

2.2 Der Bootstrap-Test

Eine weitere Möglichkeit eine Population auf Homogenität zu testen liefert der Bootstrap-Test. Dabei wird auf Grundlage von nur einer Stichprobe die asymptotische Nullverteilung der Teststatistik sowie der dazugehörige p -Wert mit der Methode des Resampling geschätzt. McLachlan (1987) verwendet für den Bootstrap-Test die Likelihood-Quotienten-Test-Statistik R_n zur Ermittlung der Anzahl der Komponenten eines Mischungsmodells. Alternativ kann auch die Teststatistik M_n des modifizierten Likelihood-Quotienten-Tests dafür herangezogen werden. Im Folgenden wird der Bootstrap-Test nur für M_n erläutert, für R_n verläuft der Test analog. Außerdem werden weiterhin nur einparametrische Mischungsmodelle mit zwei Komponenten betrachtet. Es sei X_1, \dots, X_n eine Stichprobe der Mischungsdichte

$$f_{mix}(x; \Psi^{(2)}) = (1 - \alpha)f(x; \theta_1) + \alpha f(x; \theta_2)$$

mit unbekanntem Parametern $\Psi^{(2)} = (\alpha, \theta_1, \theta_2)$ und m die Anzahl der Komponenten des Mischungsmodells. Zu testen ist also

$$H_0 : m = 1 \quad \text{gegen} \quad H_1 : m = 2.$$

Dafür werden B unabhängige sogenannte *Bootstrap-Stichproben* $X_{j1}^*, \dots, X_{jn}^*$ mit $j = 1, \dots, B$ der Dichte $f(x; \hat{\theta}_0)$ generiert, wobei $\hat{\theta}_0$ der Maximum-Likelihood-Schätzer für den Parameter der Dichte der ursprünglichen Stichprobe unter der Hypothese sei. Für jede dieser Bootstrap-Stichproben wird die Teststatistik M_n^* berechnet, anhand derer die asymptotische Nullverteilung der Teststatistik M_n der ursprünglichen Stichprobe geschätzt werden kann.

Durch den Vergleich von M_n mit den Teststatistiken der Bootstrap-Stichproben lässt sich eine Annäherung für den p -Wert von M_n bestimmen. Es seien $M_n^{*(1)}, \dots, M_n^{*(B)}$ die aufsteigend geordneten Bootstrap-Teststatistiken. Der Wert von $M_n^{*(j)}$ mit $j \in \{1, \dots, B\}$ kann als Schätzer für das Quantil $q_{\frac{j}{B}}$ betrachtet werden. Wählt man j derart, dass $M_n^{*(j-1)} \leq M_n < M_n^{*(j)}$ ist, so erhält man durch $1 - \frac{j-1}{B}$ einen Schätzer für den p -Wert. Ist dieser Wert kleiner als das vorgegebene Signifikanzniveau α , so wird die Hypothese verworfen.

Der folgende Pseudocode soll das Verfahren des Bootstrap-Tests verdeutlichen.

Pseudocode – Bootstrap-Test

gegeben: Stichprobe X_1, \dots, X_n

$P(\theta)$ parametrische Wahrscheinlichkeitsverteilung

$H_0 : X_j \sim P(\theta_0), \quad j = 1, \dots, n$

$H_1 : X_j \sim (1 - \alpha)P(\theta_1) + \alpha P(\theta_2), \quad j = 1, \dots, n$

B Anzahl der Bootstrap-Stichproben

α_{nom} nominales Niveau des Tests

berechne $\hat{\theta}_0 = \underset{\Theta}{\operatorname{argmax}} pl_n(1/2, \theta, \theta)$ und $(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2) = \underset{[0,1] \times \Theta^2}{\operatorname{argmax}} pl_n(\alpha, \theta_1, \theta_2)$

berechne $M_n = 2\{pl_n(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2) - pl_n(1/2, \hat{\theta}_0, \hat{\theta}_0)\}$

for(i in 1:B)

generiere n Zufallszahlen X_1^*, \dots, X_n^* mit $X_j^* \sim P(\hat{\theta}_0), \quad j = 1, \dots, n$

berechne $\hat{\theta}_0^* = \underset{\Theta}{\operatorname{argmax}} pl_n(1/2, \theta, \theta)$ und $(\hat{\alpha}^*, \hat{\theta}_1^*, \hat{\theta}_2^*) = \underset{[0,1] \times \Theta^2}{\operatorname{argmax}} pl_n(\alpha, \theta_1, \theta_2)$

berechne $M_n^* = 2\{pl_n(\hat{\alpha}^*, \hat{\theta}_1^*, \hat{\theta}_2^*) - pl_n(1/2, \hat{\theta}_0^*, \hat{\theta}_0^*)\}$

berechne den p -Wert: $p =$ relative Anzahl der $M_n^* > M_n$

berechne das empirische Niveau $\alpha_{emp} =$ relative Anzahl der p -Werte $\leq \alpha_{nom}$

Kapitel 3

Simulationsstudie

3.1 Likelihood-Quotienten-Test

In Satz 2.1.1 wurde gezeigt, dass die asymptotische Nullverteilung der Likelihood-Quotienten-Test-Statistik das Supremum eines Gaußprozesses ist, welches vom Parameter θ und von der Kernfunktion $f(x; \theta)$ abhängt. Diese Abhängigkeit soll im ersten Teil der Simulationsstudie für die Poisson- und Normalverteilung mit festem Erwartungswert gezeigt werden. Das Vorgehen ist für beide Kernfunktionen gleich: Unter der Hypothese wird eine Stichprobe der Größe $n = 1000$ bzw. $n = 2000$ generiert. Mittels der Maximum-Likelihood-Methode werden unter der Hypothese und der Alternative Schätzer für die Parameter der Verteilung ermittelt, mit denen die LQT-Statistik berechnet wird. Dieses Verfahren wird 20.000 mal wiederholt.

Die Abbildungen 3.1 bis 3.4 zeigen die empirischen Verteilungsfunktionen der LQT-Statistiken für verschiedene Parameter der jeweiligen Kernfunktion. In jeder der Grafiken ist außerdem die asymptotische Nullverteilung der MLQT-Statistik abgebildet. Dabei lässt sich erkennen, dass die empirischen Verteilungsfunktionen deutlich unterhalb der Verteilungsfunktion der MLQT-Statistik verlaufen, was sich damit erklären lässt, dass die MLQT-Statistik aufgrund der Penalty-Funktion stets kleiner als die LQT-Statistik ist.

Zum Vergleich werden in Abbildung 3.5 die empirischen Verteilungsfunktionen der MLQT-Statistiken für die gleichen Parameter der Poissonverteilung, die vorher auch für den LQT gewählt wurden, dargestellt.

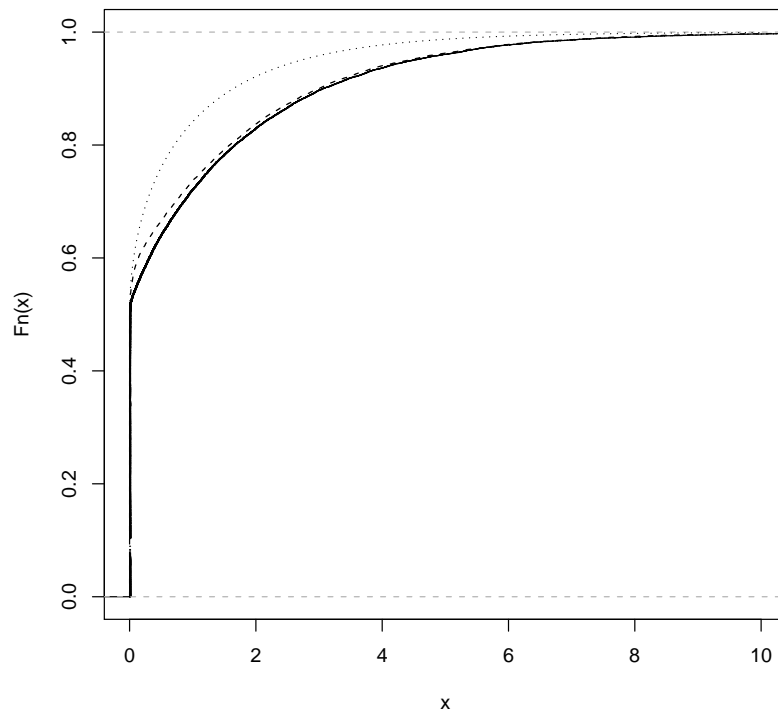


Abbildung 3.1: Kernfunktion $Poi(\lambda)$, Stichprobengröße $n = 1000$; empirische Verteilungsfunktion der LQT-Statistik für $\lambda_0 = 5$ (durchgezogener Graph), empirische Verteilungsfunktion der LQT-Statistik für $\lambda_0 = 50$ (gestrichelter Graph), asymptotische Nullverteilung der MLQT-Statistik (gepunkteter Graph)

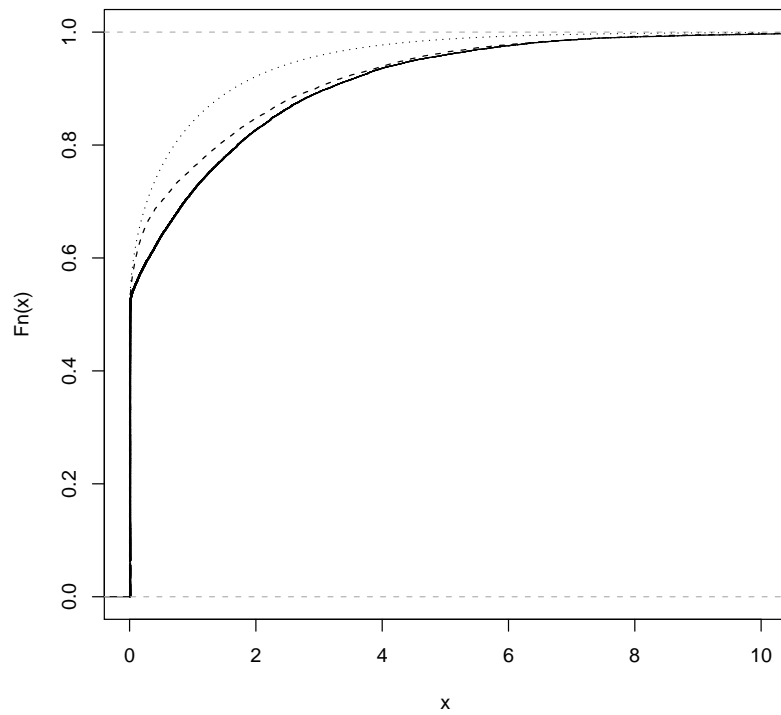


Abbildung 3.2: Kernfunktion $Poi(\lambda)$, Stichprobengröße $n = 2000$; empirische Verteilungsfunktion der LQT-Statistik für $\lambda_0 = 5$ (durchgezogener Graph), empirische Verteilungsfunktion der LQT-Statistik für $\lambda_0 = 50$ (gestrichelter Graph), asymptotische Nullverteilung der MLQT-Statistik (gepunkteter Graph)

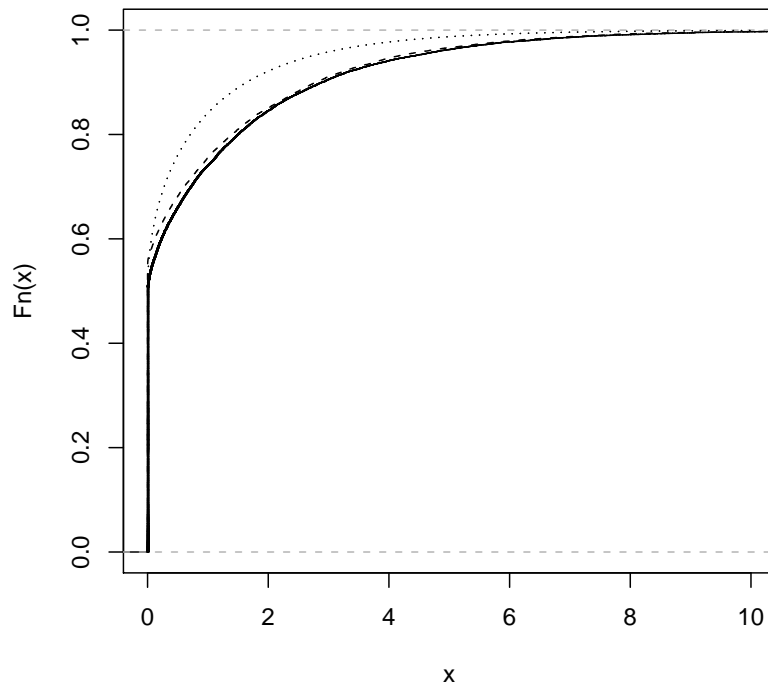


Abbildung 3.3: Kernfunktion $\mathcal{N}(0, \sigma^2)$, Stichprobengröße $n = 1000$;
empirische Verteilungsfunktion der LQT-Statistik für $\mathcal{N}(0, 1)$
(durchgezogener Graph), empirische Verteilungsfunktion der
LQT-Statistik für $\mathcal{N}(0, 20)$ (gestrichelter Graph), asymptotische
Nullverteilung der MLQT-Statistik (gepunkteter Graph)

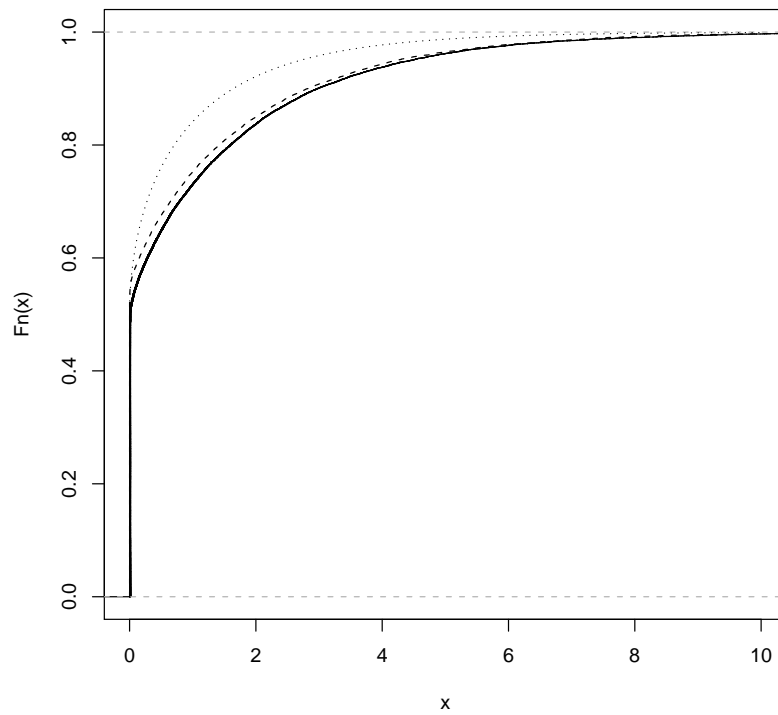


Abbildung 3.4: Kernfunktion $\mathcal{N}(0, \sigma^2)$, Stichprobengröße $n = 2000$;
empirische Verteilungsfunktion der LQT-Statistik für $\mathcal{N}(0, 1)$
(durchgezogener Graph), empirische Verteilungsfunktion der
LQT-Statistik für $\mathcal{N}(0, 20)$ (gestrichelter Graph), asymptotische
Nullverteilung der MLQT-Statistik (gepunkteter Graph)

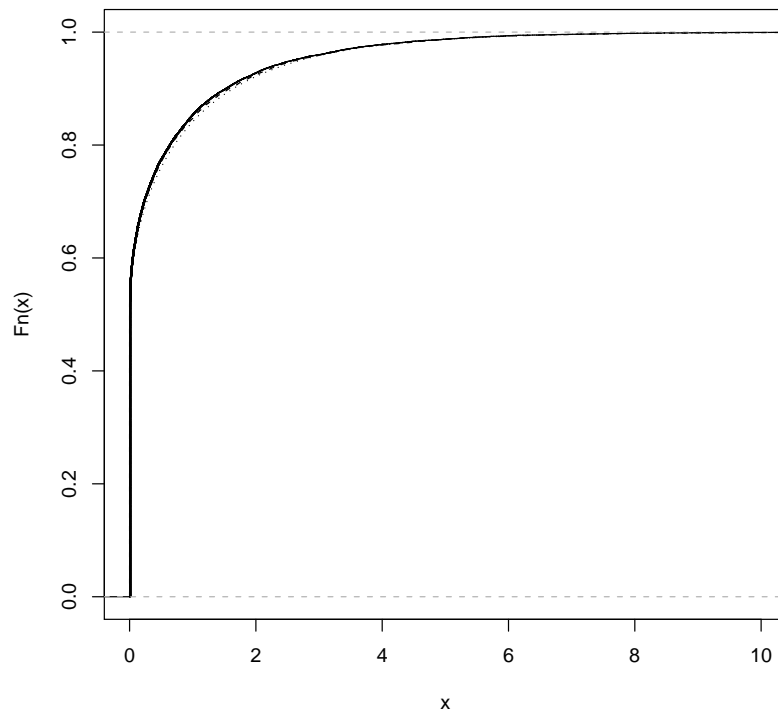


Abbildung 3.5: Kernfunktion $Poi(\lambda)$, Stichprobengröße $n = 1000$; empirische Verteilungsfunktion der MLQT-Statistik für $\lambda_0 = 5$ (durchgezogener Graph), empirische Verteilungsfunktion der MLQT-Statistik für $\lambda_0 = 50$ (gestrichelter Graph), asymptotische Nullverteilung der MLQT-Statistik (gepunkteter Graph)

3.2 Vergleich der Testniveaus des MLQT und des Bootstrap-Tests

Im Vordergrund der Simulationsstudie steht der Vergleich des modifizierten Likelihood-Quotienten-Tests mit dem Bootstrap-Test. In einem ersten Durchgang werden die empirischen Niveaus der beiden Tests betrachtet. Das Niveau α eines Tests gibt an, mit welcher Wahrscheinlichkeit die Hypothese fälschlicherweise verworfen wird. Als Kernfunktion $f(x; \theta)$ werden Poisson-, Normal- und Binomialverteilungen verwendet. Die Simulationen werden für jede der Kernfunktionen mit verschiedenen Werten für den Parameter θ_0 unter der Hypothese durchgeführt. Dabei wird bei der Normalverteilung zwischen Verteilungen aus Lokations- und Skalenfamilien unterschieden. Bei der Binomialverteilung ist die Anzahl der Bernoulliexperimente n_0 fest, die Werte für p_0 variieren (vgl. Kapitel 2.1.1).

Das Verfahren bei der Simulation des MLQT ist für jede der vier Kernfunktionen mit dem jeweils gewählten Parameter θ_0 identisch: Unter der Hypothese werden jeweils 20.000 Stichproben der Größen $n = 100, n = 200$ bzw. $n = 500$ generiert. Für jede dieser Stichproben wird der Wert der Teststatistik M_n berechnet. Unter der Hypothese besitzt die Teststatistik die asymptotische Verteilung $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Als nominale Niveaus werden $\alpha = 10\%, \alpha = 5\%$ und $\alpha = 1\%$ gewählt und für jedes Niveau das entsprechende $(1 - \alpha)$ -Quantil bestimmt. Ist der Wert der Teststatistik M_n größer als der Wert des jeweiligen Quantils, so wird die Hypothese verworfen. Auf diese Weise lässt sich die Verwerfungsrate der Hypothese bestimmen, die dem empirischen Niveau des MLQT entspricht.

Für die Simulation des Bootstrap-Tests wird eine Stichprobe der jeweiligen Verteilung unter der Hypothese vom Stichprobenumfang $n = 100$ generiert und ihre MLQT- bzw. LQT-Statistik bestimmt. Anschließend werden die entsprechenden Teststatistiken für 200 Bootstrap-Stichproben berechnet und der dazugehörige p -Wert ermittelt. Dieser Vorgang wird 5.000 mal wiederholt. Das empirische Niveau des Bootstrap-Tests ergibt sich aus der relativen Anzahl der p -Werte, die kleiner als das vorgegebene nominale Niveau sind. Die Ergebnisse dieser Simulationen sind in den Tabellen 3.1 bis 3.4 zusammengefasst¹.

¹Für mit „x“ gekennzeichnete Stellen liegen keine Ergebnisse vor.

Tabelle 3.1: Vergleich der nominalen und empirischen Niveaus (in %) des MLQT und des Bootstrap-Tests, Kernfunktion: $Poi(\lambda)$

λ_0	5			10			15		
Niveau	10	5	1	10	5	1	10	5	1
MLQT									
$n = 100$	8.63	4.42	0.96	8.66	4.46	0.95	8.36	4.34	0.99
$n = 200$	8.93	4.62	1.07	8.99	4.71	1.05	9.17	4.84	0.96
$n = 500$	9.38	4.87	1.11	9.16	4.59	0.98	9.31	4.78	0.98
Bootstrap									
$n = 100$									
MLQT	10.72	5.4	1.48	11.34	5.76	1.62	10.88	6.08	1.56
LQT	9.52	4.64	1.38	10.6	5.78	1.72	9.88	4.9	1.38

Tabelle 3.2: Vergleich der nominalen und empirischen Niveaus (in %) des MLQT und des Bootstrap-Tests, Kernfunktion: $\mathcal{N}(\mu, 1)$

μ_0	0			1			2		
Niveau	10	5	1	10	5	1	10	5	1
MLQT									
$n = 100$	8.93	4.57	0.97	8.71	4.44	0.88	8.62	4.4	0.97
$n = 200$	8.73	4.49	0.94	9.08	4.49	0.94	8.82	4.28	0.98
$n = 500$	9.64	4.54	0.97	9.43	4.79	0.99	9.27	4.78	0.96
Bootstrap									
$n = 100$									
MLQT	x	x	x	9.8	5.1	1.58	10.58	5.14	1.42
LQT	10.28	5.26	1.52	10.84	5.66	1.22	10.56	5.4	1.5

Tabelle 3.3: Vergleich der nominalen und empirischen Niveaus (in %) des MLQT und des Bootstrap-Tests, Kernfunktion: $\mathcal{N}(0, \sigma^2)$

σ_0	1			2			3		
Niveau	10	5	1	10	5	1	10	5	1
MLQT									
$n = 100$	9.34	5.04	1.29	9.21	5.13	1.26	8.7	4.71	1.17
$n = 200$	9.44	4.96	1.15	9.05	4.95	1.23	8.98	4.82	1.13
$n = 500$	9.81	5.48	1.26	9.33	4.94	1.2	9.08	4.66	1.16
Bootstrap									
$n = 100$									
MLQT	x	x	x	x	x	x	10.98	6.02	1.48
LQT	10.4	5.6	1.6	10.52	5.48	1.54	10.36	5.6	1.56

Tabelle 3.4: Vergleich der nominalen und empirischen Niveaus (in %) des MLQT und des Bootstrap-Tests, Kernfunktion: $Bin(n_0, p)$ mit $n_0 = 20$

p_0	0.3			0.5			0.7		
Niveau	10	5	1	10	5	1	10	5	1
MLQT									
$n = 100$	9.03	4.68	1.05	9.01	4.67	0.97	8.83	4.5	0.96
$n = 200$	9.16	4.63	1.02	9.13	4.5	0.98	8.9	4.4	0.94
$n = 500$	9.57	4.93	1.02	9.35	4.75	1.12	9.35	4.64	1
Bootstrap									
$n = 100$									
MLQT	10.3	5.5	1.48	10.38	5.5	1.54	10.08	5.16	1.4
LQT	11.02	5.56	1.58	10.26	5.1	1.16	10.72	5.08	1.28

3.3 Vergleich der Power des MLQT und des Bootstrap-Tests

Als nächstes wird die Power des MLQT mit der des Bootstrap-Tests verglichen. Die Power eines Tests gibt die Wahrscheinlichkeit an, mit welcher die Hypothese richtigerweise verworfen wird. Die Simulationen werden mit den gleichen Kernfunktionen wie bei der Bestimmung der empirischen Niveaus in Kapitel 3.2 durchgeführt.

Für die Simulation des MLQT wird eine Stichprobe der Größe $n = 100$ bzw. $n = 200$ unter der Alternative generiert. Die Mischungsparameter sind in Tabelle 3.5 zusammengefasst. Anschließend wird die MLQT-Statistik berechnet. Dieser Vorgang wird 10.000 mal wiederholt. Als Testniveaus werden erneut $\alpha = 10\%$, $\alpha = 5\%$ und $\alpha = 1\%$ gewählt. Die Power des MLQT ergibt sich aus der relativen Anzahl der Teststatistiken, deren Wert größer als der Wert des jeweiligen Quantils ist.

Beim Bootstrap-Test wird eine Stichprobe der Größe $n = 100$ unter der Alternative generiert und ihre MLQT- bzw. LQT-Statistik bestimmt. Anschließend werden die entsprechenden Teststatistiken für 200 Bootstrap-Stichproben berechnet und der dazugehörige p -Wert ermittelt. Dieser Vorgang wird 2.000 mal wiederholt. Die Power ergibt sich aus der relativen Anzahl der p -Werte, die kleiner als das vorgegebene nominale Niveau sind.

Die Ergebnisse der Simulationen zum Vergleich der Power der beiden Tests sind in den Tabellen 3.6 bis 3.9 dargestellt.

Tabelle 3.5: Mischungsparameter

Modell	α	θ_1	θ_2
Poisson			
I	0.5	3	6
II	0.75	3	6
III	0.9	2	5
IV	0.95	3	8
Normal mit $\sigma^2 = 1$			
I	0.5	-0.5	0.5
II	0.75	-0.9	0.3
III	0.9	-1.5	0.2
IV	0.95	-2	0.1
Normal mit $\mu = 0$			
I	0.5	0.5	1
II	0.75	1	2.5
III	0.9	1	4
IV	0.95	0.1	0.6
Binomial mit $n = 20$			
I	0.5	0.6	0.7
II	0.75	0.2	0.3
III	0.9	0.4	0.9
IV	0.95	0.2	0.9

Tabelle 3.6: Vergleich der Power (in %) des MLQT und des Bootstrap-Tests,
Kernfunktion: $Poi(\lambda)$

Level	10	5	1
MLQT			
	$n = 100$		
I	92.62	88.56	74.81
II	76.78	66.64	43.14
III	44.48	32.78	15.64
IV	43.15	32.4	16.58
	$n = 200$		
I	98.82	98.42	96.22
II	94.06	89.51	74.89
III	65.4	53.49	30.81
IV	60.59	50.05	31.33
Bootstrap (MLQT)			
	$n = 100$		
I	95.6	92.05	78.0
II	78.95	69.4	47.4
III	48.3	35.8	18.4
IV	44.85	34.6	19.25
Bootstrap (LQT)			
	$n = 100$		
I	93.75	88.1	70.75
II	76.0	65.1	43.0
III	47.6	35.55	19.0
IV	50.9	39.15	20.1

Tabelle 3.7: Vergleich der Power (in %) des MLQT und des Bootstrap-Tests,
Kernfunktion: $\mathcal{N}(\mu, 1)$

Level	10	5	1
MLQT			
	$n = 100$		
I	60.48	47.16	24.9
II	63.89	51.98	28.33
III	60.57	49.99	29.97
IV	51.52	41.18	24.17
	$n = 200$		
I	82.21	72.62	47.74
II	85.57	76.7	55.71
III	81.75	73.44	54.67
IV	71.14	62.03	44.01
Bootstrap (MLQT)			
	$n = 100$		
I	62.44	49.77	28.24
II	69.06	56.56	32.67
III	63.25	52.6	33.75
IV	52.73	41.32	25.96
Bootstrap (LQT)			
	$n = 100$		
II	63.0	49.75	27.65
III	64.4	52.9	32.7

Tabelle 3.8: Vergleich der Power (in %) des MLQT und des Bootstrap-Tests,
Kernfunktion: $\mathcal{N}(0, \sigma^2)$

Level	10	5	1
MLQT			
	$n = 100$		
I	68.91	57.19	34.11
II	57.36	45.52	23.69
III	35.08	25.55	10.99
IV	23.42	15.91	5.97
	$n = 200$		
I	88.95	82.51	63.19
II	77.58	67.89	45.76
III	48.94	38.08	20.43
IV	29.3	21.15	8.89
Bootstrap (MLQT)			
	$n = 100$		
I	70.8	57.75	35.15
II	58.3	45.1	23.9
III	33.0	23.95	11.45
IV	22.1	15.45	6.1
Bootstrap (LQT)			
	$n = 100$		
I	69.05	54.35	30.4
II	55.9	42.9	21.05
III	39.45	28.05	11.35
IV	26.45	17.95	7.75

Tabelle 3.9: Vergleich der Power (in %) des MLQT und des Bootstrap-Tests,
 Kernfunktion: $Bin(n_0, p)$ mit $n_0 = 20$

Level	10	5	1
MLQT			
	$n = 100$		
I	26.28	20.61	10.89
II	44.44	32.22	14.55
III	78.15	69.41	50.58
IV	53.37	48.48	36.75
	$n = 200$		
I	65.25	55.95	34.05
II	65.46	52.53	28.28
III	93.89	92.13	86.76
IV	79.71	75.62	65.42
Bootstrap (MLQT)			
	$n = 100$		
II	49.95	35.45	17.95
Bootstrap (LQT)			
	$n = 100$		
II	42.2	30.55	14.05

Kapitel 4

Ausblick

Die Abhängigkeit der asymptotischen Nullverteilung der LQT-Statistik vom Parameter θ wird durch die Abbildungen 3.1 bis 3.4 bestätigt. Für verschiedene Parameterwerte der gleichen Kernfunktion unterscheiden sich die empirischen Verteilungsfunktionen der Teststatistiken deutlich. Bei einem größeren Stichprobenumfang ist dieser Unterschied noch stärker zu erkennen.

Im Vergleich dazu wird in Abbildung 3.5 gezeigt, dass sich für verschiedene Parameter die empirischen Verteilungsfunktionen der MLQT-Statistiken kaum unterscheiden und nahe an der asymptotischen Nullverteilung verlaufen.

Aus den in den Tabellen 3.1 bis 3.4 zusammengefassten Ergebnissen lässt sich schließen, dass der MLQT ein eher konservativer Test ist, da das empirische Testniveau meist niedriger als das nominale Niveau ist und somit die Hypothese seltener abgelehnt wird als durch das Niveau vorgegeben. Der Bootstrap-Test hingegen lässt sich vielmehr als antikonservativ bezeichnen, da in den meisten Fällen das empirische Niveau größer als das nominale Niveau ist.

Weiterhin lässt sich feststellen, dass die Power des MLQT mit steigender Stichprobengröße wächst. Die Power des Bootstrap-Tests ist in den meisten Fällen etwas höher als die des MLQT. Dies lässt sich damit erklären, dass sich der MLQT für die in der Simulationsstudie gewählten Werte konservativ verhält, während der Bootstrap-Test eher antikonservativ ist. Somit wird die Hypothese beim Bootstrap-Test öfter abgelehnt als beim MLQT.

Bei vergleichbarer Power und Einhaltung des nominalen Niveaus ist es allerdings günstiger den MLQT zu nutzen, da der Bootstrap-Test einen viel höheren Rechenaufwand als der MLQT aufweist.

Anhang A

Die Maximum-Likelihood-Methode

Für unabhängige und identisch verteilte Zufallsvariablen X_1, \dots, X_n mit Realisierungen x_1, \dots, x_n , deren Verteilung bis auf einen unbekanntem Parameter θ bekannt ist, lässt sich mithilfe der Maximum-Likelihood-Methode ein guter Schätzer für θ ermitteln. Hierbei fasst man die Verteilung als Funktion von θ auf und erhält durch deren Maximierung denjenigen Wert für θ , für den das Eintreten der beobachteten Werte für X_1, \dots, X_n am wahrscheinlichsten ist. Aufgrund der Monotonie der Logarithmusfunktion erhält man durch die Maximierung der logarithmierten – und dadurch meist vereinfachten – Funktion von θ denselben Schätzer.

Definition A.1 Seien X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit Dichte $f(x; \theta)$, $\theta \in \Theta \subset \mathbb{R}^k$ und x_1, \dots, x_n ihre Realisierungen. Dann ist die *Likelihood-Funktion* definiert als

$$L_n(\theta) = L_n(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta).$$

Durch Logarithmieren von $L_n(\theta)$ erhält man die *Log-Likelihood-Funktion*, also

$$l_n(\theta) = \log L_n(\theta).$$

Der *Maximum-Likelihood-Schätzer* von θ ist eine Funktion $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$, sodass

$$L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta)$$

oder äquivalent

$$l_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} l_n(\theta)$$

gilt (vgl. Ferguson (1996)).

Anhang B

Der Likelihood-Quotienten-Test

Sei X_1, \dots, X_n eine unabhängige Stichprobe der Dichte $f(x; \theta)$ und sei $\Theta \subset \mathbb{R}^k$ mit $k \geq 1$ ein Parameterraum. Der Likelihood-Quotienten-Test liefert eine Methode um

$$H_0 : \theta \in \Theta_0 \quad \text{gegen} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

zu testen, wobei Θ_0 eine Teilmenge von Θ ist.

Die Nullhypothese wird verworfen, wenn die Teststatistik

$$\lambda_n = \frac{L_n(\hat{\theta}_n^*)}{L_n(\hat{\theta}_n)} = \frac{\sup_{\theta \in \Theta_0} L_n(\theta)}{\sup_{\theta \in \Theta} L_n(\theta)}$$

klein ist.

Unter den nachstehenden von Ferguson (1996) aufgestellten Regularitätsbedingungen lässt sich die in Satz B.1 dargestellte asymptotische Verteilung der Teststatistik berechnen.

Seien X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit Dichte $f(x; \theta)$ bezüglich eines dominierenden Maßes ν . Dann lauten die Regularitätsbedingungen:

- a1. Θ ist eine offene Teilmenge von \mathbb{R}^k , $k \geq 1$.
- a2. $f(x; \theta)$ ist für alle x zweimal stetig differenzierbar bezüglich θ und es gilt die Vertauschbarkeit von Integration und Differentiation:
$$\frac{\partial}{\partial \theta} \int f(x|\theta) \nu(dx) = \int \frac{\partial}{\partial \theta} f(x|\theta) \nu(dx).$$
- a3. Es existiert eine Funktion $K(x)$ mit $E_{\theta_0}[K(X)] < \infty$ und
 $|\dot{\Psi}(x, \theta)| \leq K(x)$ in einer Umgebung von θ_0 für $\dot{\Psi}(x, \theta) = \frac{\partial}{\partial \theta} \log f(x|\theta)$.

a4. Die Fisher-Information $\mathcal{I}(\theta_0) = -E_{\theta_0}[\dot{\Psi}(X, \theta_0)]$ ist positiv definit.

a5. $f(x|\theta) = f(x|\theta_0)$ ν -fast überall $\Rightarrow \theta = \theta_0$ (Identifizierbarkeit)

Satz B.1 (Ferguson, 1996) Sei $\theta^T = (\theta^1, \theta^2, \dots, \theta^k)$ ein k -dimensionaler Parameter und seien die Regularitätsbedingungen a1-a5 erfüllt. Weiterhin erfülle der wahre Wert θ_0 die Nullhypothese $H_0 : \theta^1 = \theta^2 = \dots = \theta^r = 0$ mit $1 \leq r \leq k$. Dann gilt

$$-2 \log \lambda_n \xrightarrow{d} \chi_r^2.$$

Anhang C

Sätze zur Identifizierbarkeit von endlichen Mischungen

Die hier aufgeführten Sätze werden in Kapitel 2.1.1 für die Beweise zur Identifizierbarkeit von Mischungen der verschiedenen Kernfunktionen verwendet. Mit dem folgenden Satz lässt sich die Identifizierbarkeit von endlichen Mischungen von Poisson- und Normalverteilungen zeigen.

Satz C.1 (Teicher, 1963) Sei $\mathcal{F} = \{F\}$ eine Familie von Verteilungsfunktionen mit Transformationen $\phi(t)$, $t \in S_\phi$ (S_ϕ Definitionsbereich von ϕ), sodass die Abbildung $M : F \rightarrow \phi$ linear und bijektiv ist. Es definiere \preceq eine totale Ordnung auf \mathcal{F} , sodass $F_1 \prec F_2$ impliziert, dass

(i) $S_{\phi_1} \subseteq S_{\phi_2}$,

(ii) ein $t_1 \in \bar{S}_{t_1}$ existiert (t_1 unabhängig von ϕ_2), sodass $\lim_{t \rightarrow t_1} \frac{\phi_2(t)}{\phi_1(t)} = 0$ ist.

Dann ist die Klasse aller endlichen Mischungen von \mathcal{F} identifizierbar.

Der Beweis zu diesem Satz ist nachzulesen bei Teicher (1963).

Aus den beiden nächsten Sätzen lässt sich auf die starke Identifizierbarkeit von endlichen Mischungen von Normalverteilungen aus Lokations- und Skalenfamilien schließen.

Satz C.2 (Chen, 1995) Sei $f(x)$ eine identifizierbare Dichtefunktion und $F(x, \theta) = \int_{-\infty}^x f(t - \theta) dt$. Dann erfüllt $F(x, \theta)$ die Bedingung für die starke Identifizierbarkeit, wenn $\lim_{x \rightarrow \pm\infty} f(x) = \lim_{x \rightarrow \pm\infty} f'(x) = 0$ ist.

Satz C.3 (Chen, 1995) *Aus Satz C.2 folgt die starke Identifizierbarkeit auch dann, wenn $F(x, \theta) = \frac{1}{\theta} \int_{-\infty}^x f\left(\frac{t}{\theta}\right) dt$ mit $\theta \in (0, \infty)$ ist.*

Die Beweise zu beiden Sätzen sind bei Chen (1995) zu finden.

Der nächste Satz beinhaltet eine Bedingung für die Identifizierbarkeit von endlichen Mischungen von Binomialverteilungen.

Satz C.4 (Teicher, 1963) *Sei $\mathcal{F} = \{F(x; n_0, p), 0 < p < 1, n_0 \in \mathbb{N} \text{ fest}\}$ eine einparametrische Familie von Binomialverteilungen. Die Klasse aller endlichen Mischungen aus maximal k Komponenten aus \mathcal{F} ist genau dann identifizierbar, wenn*

$$n_0 \geq 2k - 1.$$

Den Beweis zu diesem Satz liefern Titterton et al. (1985).

Im folgenden Satz stellen Yakowitz und Spragins (1968) eine Möglichkeit vor, um zu zeigen, dass eine parametrische Familie von Verteilungsfunktionen identifizierbar ist.

Satz C.5 (Yakowitz und Spragins, 1968) *Eine notwendige und hinreichende Bedingung für die Identifizierbarkeit von Mischungen einer Familie von Verteilungsfunktionen $\mathcal{F} = \{F(x; \theta), x \in \mathbb{R}, \theta \in \Theta\}$ ist die lineare Unabhängigkeit der Komponenten von \mathcal{F} für alle $x \in \mathbb{R}$.*

Der Beweis ist bei Yakowitz und Spragins (1968) nachzulesen.

Anhang D

Wahrscheinlichkeits- und momentenerzeugende Funktionen

In den Beweisen zur Identifizierbarkeit von endlichen Mischungen von Poisson-, Normal- und Binomialverteilungen in Kapitel 2.1.1 werden wahrscheinlichkeits- und momentenerzeugende Funktionen verwendet. Die Definitionen für solche Funktionen werden hier aufgeführt.

Definition D.1 Sei X eine \mathbb{N}_0 -wertige Zufallsvariable. Die Abbildung

$$g : [0, 1] \rightarrow [0, 1], \quad z \mapsto \sum_{n=0}^{\infty} \mathbf{P}[X = n]z^n,$$

heißt wahrscheinlichkeitserzeugende Funktion von \mathbf{P}_X (vgl. Klenke (2008)).

Beispiel D.2 (i) Sei X $b_{n,p}$ -verteilt für gewisse $n \in \mathbb{N}$ und $p \in [0, 1]$.

Dann ist

$$g(z) = \sum_{m=0}^n \binom{n}{m} p^m (1-p)^{n-m} z^m = (pz + (1-p))^n.$$

(ii) Sei X poissonverteilt mit Parameter $\lambda \geq 0$. Dann ist

$$g(z) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda z)^n}{n!} = e^{\lambda(z-1)}$$

(vgl. Klenke (2008)).

Definition D.3 Ist X eine reelle Zufallsvariable und $D := \{s \in \mathbb{R} : E[\exp(sX)] < \infty\}$, so heißt die Funktion

$$m : D \rightarrow \mathbb{R}, \quad s \mapsto E[\exp(sX)] = \int \exp(sx) d\mathbf{P}_X(x)$$

momentenerzeugende Funktion (vgl. Meintrup und Schäffler (2005)).

Beispiel D.4 (i) Sei X eine poissonverteilte Zufallsvariable mit Parameter $\lambda \geq 0$. Dann ist

$$m(s) = \sum_{n=0}^{\infty} e^{sn} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^s)^n}{n!} = e^{\lambda(e^s - 1)}.$$

(ii) Sei X eine normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2 . Dann ist

$$m(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}$$

(vgl. Meintrup und Schäffler (2005) und Hesse (2009)).

In den Beweisen zur Identifizierbarkeit von endlichen Mischungen von Normalverteilungen aus Lokations- und aus Skalenfamilien wird die zweiseitige Laplace-Transformation verwendet. Diese steht in engem Zusammenhang zur momentenerzeugenden Funktion.

Definition D.5 Ein endliches Maß μ auf $[0, \infty)$ ist eindeutig bestimmt durch Angabe der Laplace-Transformierten

$$\mathcal{L}_\mu(t) = \int e^{-tx} \mu(dx), \quad \text{für } t \geq 0$$

(vgl. Klenke (2008)).

Beispiel D.6 Sei X eine normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2 . Dann ist

$$\mathcal{L}_\mu(t) = e^{\frac{\sigma^2 t^2}{2} - \mu t}$$

die zweiseitige Laplace-Transformation der Verteilung (bei der zweiseitigen Laplace-Transformation wird von $-\infty$ bis ∞ integriert)(vgl. Teicher (1963)).

Literaturverzeichnis

- Chen, J. (1995). Optimal rate of convergence for finite mixture models. *The Annals of Statistics* **23**, No. 1, 221-233.
- Chen, J., Chen, H. and Kalbfleisch, D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society. Series B*, **63**, Part 1, 19-29.
- Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.
- Hesse, C. (2009). *Wahrscheinlichkeitstheorie*. Vieweg + Teubner.
- Klenke, A. (2008). *Wahrscheinlichkeitstheorie*. Springer.
- Li, P., Chen, J., Marriott, P. (2008). Non-finite Fisher information and homogeneity: an EM approach. *Biometrika Trust* **96**, No. 2, 411-426.
- McLachlan, G.J (1987). On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in an Normal Mixture. *Royal Statistical Society* **36**, No. 3, 318-324.
- McLachlan, G.J., Peel, D. (2000). *Finite Mixture Models*. Wiley.
- Meintrup, D. und Schäffler, S. (2005). *Stochastik: Theorie und Anwendungen*. Springer.
- Schlattmann, P. (2009). *Medical Applications of Finite Mixture Models*. Springer.
- Teicher, H. (1963). Identifiability of Finite Mixtures. *The Annals of Ma-*

thematical Statistics **34**, No. 4, 1265-1269.

Titterington, D.M., Smith, A.F.M., Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley.

Yakowitz, S.J. and Spragins, J.D. (1968). On the Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics* **39**, No. 1, 209-214.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen meiner Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, den 23. September 2010
