

# Verzerrung der Inferenz bei der Verwendung gemischter Modelle in latent Repräsentationen

**Yannick Bantel**

Wissenschaftliche Arbeit zur Erlangung des Grades

**Bachelor of Science**

an der TUM School of Computation, Information and Technology der Technischen Universität München

**Prüfer(in):**

Prof. Dr. Harald Binder

**Betreuer(in):**

Clemens Schächter

**Eingereicht:**

München, den 23. Juli 202



Ich versichere hiermit, dass ich die von mir eingereichte Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 23. Juli 202

Yannick Bantel



## **Zusammenfassung**

Eine kurze Zusammenfassung der Arbeit auf Deutsch.

## **Abstract**

A brief abstract of this thesis in English.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.0.1	Motivation . . . . .	1
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>3</b>
2.1	Gemischte Modelle . . . . .	3
2.2	Likelihood Inferenz . . . . .	4
2.2.1	Likelihood Berechnung gemischter Modelle . . . . .	4
2.2.2	Likelihood-Ratio-Test . . . . .	6
2.3	Variational Autoencoder . . . . .	6
2.3.1	Training VAE . . . . .	6
2.4	Verzerrung (Bias) und ihre Messung . . . . .	8
<b>3</b>	<b>Minimierung der Verzerrung</b>	<b>9</b>
<b>4</b>	<b>Methodik</b>	<b>11</b>
4.1	Vorgehen . . . . .	11
4.2	Datenbeschaffung . . . . .	11
4.3	Modellierungstechniken . . . . .	11
4.4	Analysemethoden . . . . .	11
<b>5</b>	<b>Experimente und Ergebnisse</b>	<b>13</b>
5.1	Experimentelles Design . . . . .	13
5.2	Durchführung . . . . .	13
5.3	Analyse der Ergebnisse . . . . .	13
<b>6</b>	<b>Diskussion</b>	<b>15</b>
6.1	Interpretation der Ergebnisse . . . . .	15
6.2	Vergleich mit bestehenden Arbeiten . . . . .	15
6.3	Limitationen und Herausforderungen . . . . .	15
<b>7</b>	<b>Fazit</b>	<b>17</b>
<b>8</b>	<b>Anhang</b>	<b>19</b>
<b>9</b>	<b>Literaturverzeichnis</b>	<b>21</b>
<b>A</b>	<b>Appendix</b>	<b>23</b>
A.1	Supporting Data . . . . .	23
A.2	Some Code Listings . . . . .	23
	<b>Literatur</b>	<b>29</b>





# 1 Einleitung

In der modernen Datenanalyse spielen gemischte Modelle eine zentrale Rolle, da sie es ermöglichen, sowohl feste als auch zufällige Effekte zu berücksichtigen, was sie besonders in den Bereichen der Biostatistik, der Sozialwissenschaften und der ökonomischen Modellierung populär macht. Im Rahmen dieser Arbeit werden gemischte Modelle auf die Analyse medizinischer Daten angewendet. Mit dem Aufkommen von Big Data und komplexen Datenstrukturen hat sich der Fokus zunehmend auf die effiziente und genaue Extraktion von Informationen aus großen und oft unübersichtlichen Datensätzen verlagert. In diesem Zusammenhang gewinnen latente Repräsentationen an Bedeutung, da sie es ermöglichen, inhärente Strukturen innerhalb der Daten zu identifizieren und zu nutzen, um tiefere Einblicke zu gewinnen.

Die Integration von gemischten Modellen in latente Repräsentationen birgt jedoch das Risiko einer Verzerrung der Inferenzergebnisse, was die Genauigkeit und Zuverlässigkeit der aus den Daten gezogenen Schlussfolgerungen erheblich beeinträchtigen kann. Diese Arbeit befasst sich daher mit der Untersuchung der Verzerrungen, die bei der Anwendung gemischter Modelle auf latente Repräsentationen auftreten können. Ziel ist es, die Mechanismen zu verstehen, die zu diesen Verzerrungen führen, und Methoden zu entwickeln, um ihre Auswirkungen zu minimieren.

Das Problem der Verzerrung ist besonders relevant, da eine fehlerhafte Inferenz zu Fehlentscheidungen führen kann, die in praktischen Anwendungen schwerwiegende Folgen haben können. Durch eine sorgfältige Analyse und Bewertung von gemischten Modellansätzen in Verbindung mit latenten Repräsentationen zielt diese Arbeit darauf ab, einen Beitrag zur Verbesserung der Modellgenauigkeit und der Zuverlässigkeit von Inferenzschlüssen zu leisten.

Die Arbeit gliedert sich in mehrere Teile, die zunächst die theoretischen Grundlagen von gemischten Modellen und latenten Repräsentationen behandeln, gefolgt von einer Diskussion der Methoden zur Messung und Korrektur von Verzerrungen. Anhand von experimentellen Studien werden diese Konzepte dann praktisch angewendet und evaluiert, um abschließend Empfehlungen für die Anwendung dieser Techniken in Forschung und Praxis zu geben.

## 1.0.1 Motivation



## 2 Theoretische Grundlagen

Bevor auf die Methodik dieser Arbeit eingegangen wird, ist eine theoretische Aufarbeitung der behandelten Themen notwendig. In diesem Kapitel werden die theoretischen Aspekte dieser Arbeit ausführlich beschrieben und behandelt. Sowohl lineare gemischte Modelle als auch die Theorie hinter Variational Autoencodern werden eingeführt und beschrieben. Insbesondere wird in diesem Kapitel auf die für die Analyse der Modelle notwendige Theorie, wie z.B. die Likelihood-Berechnung und den Likelihood-Ratio-Test, eingegangen.

### 2.1 Gemischte Modelle

Um die theoretischen Aspekte der gemischten Modelle zu behandeln, folgen wir dem Buch Mixed-Effects Models in S and S-Plus von José C. Pinheiro und Douglas M. Bates [PB00].

Ein gemischtes Modell ist ein statistisches Verfahren zur Datenanalyse, das sowohl feste als auch zufällige Effekte (fixed and random effects) modelliert. Gemischte Modelle werden vor allem bei der Analyse von Längsschnitt- und Clusterdaten eingesetzt. Die parametrisierten festen und zufälligen Effekte berechnen zusammen mit einem Fehlervektor die Reaktionsvariable. Das lineare gemischte Modell für eine Gruppe ist wie folgt definiert

**Definition 2.1.1** (Lineares gemischtes Modell für Longitudinal- oder Clusterdaten).

Seien  $X (n_i \times p)$  und  $Z (n_i \times q)$  bekannte Designmatrizen für die festen und zufälligen Effekte. Seien  $\beta$  ein  $p$ -dimensionaler Vektor von  $p$  festen Effekten und  $b$  ein  $q$ -dimensionaler Vektor von  $q$  zufälligen Effekten und sei  $\epsilon$  ein  $n_i$ -dimensionaler normalverteilter Fehlervektor.

Ein lineares gemischtes Modell für den  $n_i$ -dimensionalen Antwortvektor der  $i$ -ten Gruppe wird durch

$$y_i = X_i * \beta + Z_i * b_i + \epsilon_i$$

$$b_i \sim \mathcal{N}(0, \Sigma), \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)$$

definiert.

Die Daten der zufälligen und festen Effekte werden in einer Designmatrix (Datenmatrix) gespeichert. Die Parametervektoren  $\beta$  (für die festen Effekte) und  $b_i$  (für die zufälligen Effekte) initialisieren den Einfluss der Daten auf den Antwortvektor. Für immer auftretende Messfehler oder unerwartete Einflüsse wird ein zufälliges Rauschen  $\epsilon$  hinzugefügt.

Die zufälligen Effekte  $b_i$  und der Fehlervektor sind in der gleichen Gruppe unabhängig voneinander und unabhängig von anderen Gruppen. Wie man leicht erkennen kann, ist der Vektor der zufälligen Effekte nur durch seine Varianz-Kovarianz-Matrix  $\Sigma$  charakterisiert, die symmetrisch und positiv semidefinit ist. Die einzelnen Cluster/Gruppen können zu einem einzigen allgemeinen linearen gemischten Modell zusammengefasst werden:

**Definition 2.1.2** (Allgemeines lineares gemischtes Modell).

Ein lineares gemischtes Modell ist definiert durch

$$y = X\beta + Zb + \epsilon$$

mit

$$\begin{pmatrix} b \\ \epsilon \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \sigma^2 I \end{pmatrix} \right)$$

gegeben. Dabei sind  $X$ , bzw  $Z$  die Designmatrizen der festen, bzw zufälligen Effekte,  $\beta$  und  $b$  die Parametervektoren der festen und der zufälligen Effekten und  $\epsilon$  der Fehlervektor.

Die Vektoren  $b$  und  $\epsilon$  sind also normalverteilt mit der Varianz  $\text{Var}(b) = \Sigma$  und  $\text{Var}(\epsilon) = R = \sigma^2 I$ . Die Varianz von  $y$  kann daher geschrieben werden als  $\text{Var}(y) = V = Z \Sigma Z^t + R$ .

## 2.2 Likelihood Inferenz

Um den Einfluss eines fixen Effekts zu testen, vergleichen wir ein reduziertes Modell ohne diesen fixen Parameter mit dem vollständigen Modell. Dazu verwenden wir den Likelihood-Ratio-Test (LRT), der üblicherweise für diesen Zweck verwendet wird. Wie das Testverfahren genau funktioniert und wie der LRT durchgeführt wird, werden wir später sehen. Zuvor benötigen wir noch etwas Theorie zur Likelihood-Berechnung.

### 2.2.1 Likelihood Berechnung gemischter Modelle

Dieser Abschnitt befasst sich mit der Schätzung der unbekannten Parameter. Wir folgen hier dem Ansatz von [fahrmeir2010].

Wir verwenden die Maximum-Likelihood (ML)-Methode zur Berechnung der Schätzer. Eine Alternative wäre die restringierte ML-Methode, die jedoch nicht für den Likelihood-Ratio-Test geeignet ist. Daher verwenden wir für unsere Berechnung die ML-Methode.

Die Schätzung der Parameter in einem gemischten Modell ist jedoch etwas komplizierter. Nicht nur  $\beta$  ist unbekannt, sondern auch  $b$ ,  $\Sigma$  und  $R$ . Daher müssen sowohl die festen und zufälligen Effekte als auch die unbekannten Parameter, die wir als  $\theta$  bezeichnen, in  $\Sigma$  und  $R$  geschätzt werden. Dies zwingt uns zu einer geschachtelten Schätzung.

Nehmen wir zunächst an, dass  $X$ ,  $V$  und  $Z$  bekannt sind. Für die Schätzung von  $\beta$  aus dem Randmodell bietet sich der ML-Schätzer

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y$$

an. Setzt man hier  $\hat{V}$  ein, so erhält man den besten linearen erwartungstreuen Schätzer (BLUE, best linear unbiased estimator) für die fixen Effekte.

Für den Schätzer von  $b$  verwenden wir den bedingten Erwartungswert  $E(b|y)$  von  $b$ , gegeben die Daten  $y$ . Betrachten wir die gemeinsame Verteilung von  $b$  und  $y$

$$\begin{pmatrix} y \\ b \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V & Z \Sigma \\ \Sigma Z^t & \Sigma \end{pmatrix} \right)$$

erhalten wir  $E(b|y) = \Sigma Z^t V^{-1} (y - X\beta)$ .

Ersetzt man nun  $\beta$  durch den Schätzer  $\hat{\beta}$  erhält man den Schätzer

$$\hat{b} = \Sigma Z^t \hat{V}^{-1} (y - X\hat{\beta})$$

für die zufälligen Effekte. Der Schätzer  $\hat{b}$  ist der beste lineare unverzerzte Schätzer (BLUP, best linear unbiased prediction)

**Definition 2.2.1** (Schätzer für feste und zufällige Effekte).

Sei  $y = X\beta + Zb + \epsilon$  ein lineares gemischtes Modell und sei  $V = \text{Var}(y)$ . Dann ist

$$\hat{\beta} = (X^t \hat{V}^{-1} X)^{-1} X^t \hat{V}^{-1} y$$

ein Schätzer für die festen Effekte und

$$\hat{b} = \sum Z^t \hat{V}^{-1} (y - X\hat{\beta})$$

ein Schätzer für die zufälligen Effekte.

Wie bereits erwähnt, soll  $\theta$  der Parametervektor sein, der alle unbekannten Parameter in  $V = V(\theta)$ ,  $\Sigma = \Sigma(\theta)$  und  $R = R(\theta)$  enthält. Wenn wir nun einen Schätzer  $\hat{\theta}$  haben, können wir durch Einsetzen von  $\hat{\theta}$  die Kovarianzschätzer und damit die Schätzer der festen und zufälligen Effekte berechnen. Die ML-Methode für  $\theta$  basiert auf dem Randmodell

$$y \sim N(X\beta, V(\theta)).$$

Die Log-Likelihood von  $\beta$  und  $\theta$  ist gegeben durch

$$l(\beta, \theta) = -\frac{1}{2}(\log(|V|) + (y - X\beta)^t V^{-1} (y - X\beta)).$$

Maximieren von  $l(\beta, \theta)$  bezüglich  $\beta$  für festes  $\theta$  ergibt

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y.$$

Setzt man nun  $\hat{\theta}$  in  $l(\beta, \theta)$  ein, so erhält man die Profil-Log-Wahrscheinlichkeit

$$l(\theta)_p = -\frac{1}{2}(\log(|V|) + (y - X\hat{\beta})^t V^{-1} (y - X\hat{\beta})).$$

Folglich erhält man den ML-Schätzer  $\hat{\theta}_{ML}$  durch Maximierung von  $l(\theta)_p$ .

**Definition 2.2.2** (Kovarianz-Schätzer).

Sei  $y = X\beta + Zb + \epsilon$  ein lineares gemischtes Modell mit

$$\begin{pmatrix} b \\ \epsilon \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & R = \sigma^2 I \end{pmatrix} \right)$$

und sei  $\theta$  der unbekannte Parametervektor von  $\Sigma, R$  und  $V = \text{Var}(y)$ .

Dann ist  $\hat{\theta}_{ML}$  der ML-Schätzer für  $\theta$ , den man durch maximieren von

$$l(\theta)_p = -\frac{1}{2}(\log(|V|) + (y - X\hat{\beta})^t V^{-1} (y - X\hat{\beta}))$$

erhält. Dabei ist

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y$$

Mit dem Schätzer  $\hat{V}$  lassen sich die Schätzer der festen und zufälligen Effekte nun berechnen.

**Definition 2.2.3** (Maximum-Likelihood).

Sei

$$l(\Sigma, R) = -0.5 * (\log(|V|) + r' V^{-1} r + N * \log(2\pi))$$

### 2.2.2 Likelihood-Ratio-Test

Die Berechnung der Likelihood-Ratio-Test (LRT)-Statistik ist relativ einfach, wenn man sich die Theorie der ML-Methode vergegenwärtigt. Zur Erinnerung: Wir wollen ein reduziertes Modell mit dem vollständigen Modell vergleichen, um herauszufinden, wie groß der Einfluss einer Störgröße ist. Dazu verwenden wir den LRT.

**Definition 2.2.4** (Likelihood-Ratio-Test (LRT)).

Sei  $\mathbf{L}_{full}$  der Likelihood-Wert des vollständigen Modells und  $\mathbf{L}_{red}$  der Likelihood-Wert des reduzierten Modells. Sei  $i$  die Anzahl der Freiheitsgrade. Dann ist die LRT Statistik gegeben durch

$$LRT = 2(\mathbf{L}_{full} - \mathbf{L}_{red})$$

Falls  $\mathbf{L}_{full}$  und  $\mathbf{L}_{red}$  wie in der Definition initialisiert sind, gilt  $\mathbf{L}_{full} > \mathbf{L}_{red}$ . Insbesondere gilt  $\log(\mathbf{L}_{full}) > \log(\mathbf{L}_{red})$ . Falls  $\mathbf{L}_{full}$  und  $\mathbf{L}_{red}$  nun schon die Log-Likelihood-Werte der Modelle sind, gilt für die Berechnung der LRT Statistik einfach  $2(\mathbf{L}_{full} - \mathbf{L}_{red})$ .

## 2.3 Variational Autoencoder

Für die Anwendung der gemischten Modelle in einer latenten Repräsentation wird ein Variational Autoencoder (VAE) verwendet, der für die Methodik dieser Arbeit von Vorteil ist. Ein VAE ist ein künstliches neuronales Netzwerk, bestehend aus einem Encoder  $q_\phi(z|x)$  und einem Decoder  $p_\theta(x|z)$ , das lernt, die wesentlichen Merkmale der Eingabedaten zu extrahieren und eine komprimierte Version dieser Daten zu erzeugen. Sie sind für latente Repräsentationen sehr interessant, da sie hochdimensionale Datensätze mit Hilfe des Encoders im latenten Raum niedrigdimensional darstellen können. Dies reduziert die Komplexität der Modellierung und ermöglicht es, gemischte Modelle effizienter und genauer zu betreiben. In einem niedrigdimensionalen Raum ist die Wahrscheinlichkeit geringer, dass die Modelle übermäßig komplexe oder verzerrte Strukturen lernen.

Im Vergleich zu einem üblichen Autoencoder, welcher aus dem Encoder im latenten Raum eine Vektordarstellung der Daten liefert, gibt der Encoder des VAE eine dimensionalreduzierte Darstellung der Daten als Wahrscheinlichkeitsverteilung aus. Der Encoder liefert also zwei Vektoren: den Mittelwert  $\mu$  und die Standardabweichung  $\sigma^2$ . Durch diese verbesserte Darstellung ist der VAE im Gegensatz zu normalen Auto-Encodern in der Lage nicht nur den Input-Datensatz zu rekonstruieren, sondern auch andere ähnliche Datensätze.

### 2.3.1 Training VAE

Ein VAE ist ein spezifisches Beispiel einer Variational Inference (VI). Ziel einer VI ist es die Posterior-Verteilung  $p(z|x)$  der latenten Variablen gegeben der Input Daten zu berechnen. Diese Verteilung ist besonders bei komplexen Modellen schwer zu berechnen. Aufgrund dessen approximiert der VAE die Posterior-Verteilung durch eine einfachere Verteilung  $q_\phi(x)$ . Das Ziel von VIs ist es also die Kullback-Leibler-Divergenz zwischen  $p(z|x)$  und  $q_\phi(x)$  zu minimieren.

**Definition 2.3.1** (Kullback-Leibler-Divergenz (KL-Divergenz)).

Sei  $q_\phi(z|x)$  die approximierte Posterior-Gauß-Verteilung und  $p(z)$  die Prior-Gauß-Verteilung. Dann ist die KL-Divergenz definiert als

$$D_{KL}(q_\phi(z|x)||p(z)) = \int q_\phi(z|x) \log \left( \frac{q_\phi(z|x)}{p(z)} \right) dz$$

Da nun  $p(z)$  schwer zu berechnen ist optimieren wir stattdessen die Evidence Lower Bound (ELBO).

**Definition 2.3.2** (Evidence Lower Bound (ELBO) für VAEs).

Sei  $q_\phi(z|x)$  das Encoder Modell und  $p_\theta(x, z)$  das Decoder Modell. Der ELBO ist definiert durch

$$\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)]$$

**Definition 2.3.3** (Jensensche Ungleichung).

Sei  $g$  eine konvexe Funktion und  $x \in \mathcal{L}^\infty$ . Dann gilt

$$\mathbb{E}[g(x)] \geq g(\mathbb{E}[x])$$

Aus der Definition folgt direkt  $\mathbb{E}[\log(x)] \geq \log(\mathbb{E}[x])$ .

Das Objekt der Optimierung im VAE wird also der ELBO sein.

Wir betrachten nun die log-likelihood der Daten.

$$\log p_\theta(x) = \log p_\theta(x) * \overbrace{\int q_\phi(z|x) dz}^{=1} \quad (2.1)$$

$$= \int \log p_\theta(x) q_\phi(z|x) dz \quad (2.2)$$

$$= \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x)] \quad (2.3)$$

$$= \mathbb{E}_{q_\phi(z|x)} \left[ \log \left[ \frac{p_\theta(x, z)}{p_\theta(z|x)} \right] \right] \quad (2.4)$$

$$= \mathbb{E}_{q_\phi(z|x)} \left[ \log \left[ \frac{p_\theta(x, z) q_\phi(z|x)}{q_\phi(z|x) p_\theta(z|x)} \right] \right] \quad (2.5)$$

$$= \underbrace{\mathbb{E}_{q_\phi(z|x)} \left[ \log \left[ \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] \right]}_{= \mathcal{L}_{\theta, \phi}(x) \text{ (ELBO)}} + \underbrace{\mathbb{E}_{q_\phi(z|x)} \left[ \log \left[ \frac{q_\phi(z|x)}{p_\theta(z|x)} \right] \right]}_{= D_{KL}(q_\phi(z|x) || p_\theta(z|x))} \quad (2.6)$$

Der zweite Term in der Gleichung ist die nicht-negative Kullback-Leibler-Divergenz (KL-Divergenz) zwischen  $q_\phi(z|x)$  und  $p_\theta(z|x)$ . Die KL-Divergenz misst den Unterschied, bzw. die Ähnlichkeit, zweier Wahrscheinlichkeitsverteilungen. Der erste Term in Gleichung (2.4) ist der variational lower bound, also der ELBO:

$$\mathcal{L}_{\theta, \phi}(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)]$$

Wenn wir Gleichung (2.4) umstellen sehen wir, dass der ELBO eine untere Schranke für die log-likelihood der Daten ist, da die KL-Divergenz nicht negativ ist.

$$\mathcal{L}_{\theta, \phi}(x) = \log p_\theta(x) - D_{KL}(q_\phi(z|x) || p_\theta(z|x)) \quad (2.7)$$

$$\leq \log p_\theta(x) \quad (2.8)$$

Alternativ mit der Jensenschen Ungleichung:

$$\log p_\theta(x) = \log \int p_\theta(x, z) dz \quad (2.9)$$

$$= \log \int p_\theta(x, z) \frac{q_\phi(z|x)}{q_\phi(z|x)} dz \quad (2.10)$$

$$= \log \mathbb{E}_{q_\phi(z|x)} \left[ \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] \quad (2.11)$$

$$\geq \mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right] = \mathcal{L}_{\theta, \phi}(x) \quad (2.12)$$

Es ist sofort ersichtlich, dass die log-likelihood  $p_\theta(x)$  durch Maximieren der ELBO bzgl.  $\theta$  und  $\phi$  selbst maximiert wird. Unser generatives Modell wird also besser. Gleichzeitig minimieren wir dadurch die KL-Divergenz der Approximation  $q_\phi(z|x)$  von dem wahren Vorgänger  $p_\theta(z|x)$ . Demnach wird  $q_\phi(z|x)$  optimiert. Die ELBO kann durch stochastische Gradientenverfahren optimiert werden.

Wir wollen also den Gradienten von  $\mathcal{L}_{\theta,\phi}(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x, z) - \log q_\phi(z|x)]$  bezüglich  $\theta$  berechnen, was problematisch sein kann. Der übliche Monte Carlo Gradienten-Schätzer ist:

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} [f(z)] = \mathbb{E}_{q_\phi(z)} [f(z) \nabla_{q_\phi(z)} \log q_\phi(z)] \approx \frac{1}{L} \sum_{l=1}^L f(z) \nabla_{q_\phi(z^{(l)})} \log q_\phi(z^{(l)}) \quad (2.13)$$

wobei  $z^{(l)} \sim q_\phi(z|x^{(i)})$  ist. Dies ist im Normalfall keine Herausforderung. Allerdings ist  $q_\phi(z|x)$  nun auch abhängig von  $\phi$ , sodass der Monte Carlo Gradienten-Schätzer nicht direkt berechnet werden kann.

Um dieses Problem zu lösen nutzen wir den Reparameterization Trick, welcher die Zufallsvariable transformiert um die Gradienten-Berechnung zu vereinfachen.

### Reparametrisierungs Trick

Der Reparameterisierungs-Trick ermöglicht die Berechnung der Gradienten der ELBO durch eine Transformation der Zufallsvariable  $z$  in eine deterministische Funktion von einer Hilfsvariablen  $\epsilon$ . Sei also die latente Variable  $z$ , die aus  $q_\phi(z|x)$  gezogen wurde, gegeben. Die latente Variable  $z$  wird nun als deterministische Funktion einer Hilfsvariablen  $\epsilon$ , unabhängig von  $x$  und  $\phi$ , ausgedrückt. Die Transformation sieht dann wie folgt aus

$$z = g_\phi(\epsilon, x)$$

$g_\phi(\epsilon, x)$  ist dabei eine differenzierbare Funktion und  $\epsilon$  eine Zufallsvariable mit einer bekannten Verteilung (z.B.  $\epsilon \sim \mathcal{N}(0, I)$ ) Im Falle einer Gaußverteilung  $z \sim \mathcal{N}(\mu, \sigma^2)$  könnte die Umparametrisierung wie folgt aussehen

$$z = \mu + \sigma \odot \epsilon \quad \text{mit } \epsilon \sim \mathcal{N}(0, I).$$

Dadurch können die Gradienten bezüglich  $\mu$  und  $\sigma$  effizient berechnet werden.

Wir können den Trick direkt auf Gleichung (2.13) anwenden. Es gilt

$$\nabla_\phi \mathbb{E}_{q_\phi(z)} [f(z)] = \nabla_\phi \mathbb{E}_{p(\epsilon)} [g_\phi(\epsilon, x)]$$

## 2.4 Verzerrung (Bias) und ihre Messung

Erklärung von Verzerrung, wie sie entsteht und wie sie gemessen wird.



### **3 Minimierung der Verzerrung**



## 4 Methodik

### 4.1 Vorgehen

In den ersten Wochen habe ich mir selbst ein Simulationsdesign für einen longitudinalen medizinischen Datensatz ausgedacht und basierend darauf ein gemischtes Modell gefittet. Mit diesen simulierten Daten habe ich ein reduziertes Modell mit dem vollständigen Modell verglichen. Die LRT Statistik habe ich dann in einem Histogramm dargestellt.

Wir fügen dem gemischten Modell einen festen Effekt hinzu, welcher keinen Einfluss auf die Trajektorie haben soll. In unserem Fall ist dieser feste Effekt das Geschlecht, welches keinen Einfluss auf den Verlauf einer Krankheit haben sollte.

Mein zweites Projekt ist nun einen hoch dimensionalen medizinischen Datensatz durch den Encoder eines Variational Autoencoders im latenten Raum zu repräsentieren und dort mit einem gemischten Modell darzustellen. Ähnlich wie zuvor will ich wieder eine LRT Statistik erhalten, in dem ich ein reduziertes Modell mit dem vollständigen Modell vergleiche. Dazu trainiere ich in einer Schleife den Encoder und das gemischte Modell für jeden Iterationsschritt neu und vergleiche die negativen Maximum Likelihood-Werte (ML-Werte) durch den Likelihood Ratio Test. Am Ende der Schleife erhalte ich wieder eine LRT Statistik, welche durch ein Histogramm dargestellt wird. Im Optimalfall ähnelt das Histogramm einer Chi-Quadrat-Verteilung mit einem Freiheitsgrad (Da das reduzierte Modell nur einen festen Effekt, das Geschlecht, weniger hat).

### 4.2 Datenbeschaffung

Quellen und Typen der verwendeten Daten

### 4.3 Modellierungstechniken

Beschreibung der spezifischen gemischten Modelle und der Techniken zur Gewinnung latenter Repräsentationen.

### 4.4 Analysemethoden

Verfahren zur Untersuchung der Verzerrung in den Inferenzergebnissen.



# **5 Experimente und Ergebnisse**

## **5.1 Experimentelles Design**

Aufbau der experimentellen Tests und Simulationen.

## **5.2 Durchführung**

Beschreibung der durchgeführten Experimente und verwendeten Parameter.

## **5.3 Analyse der Ergebnisse**

Diskussion der Ergebnisse im Hinblick auf die Verzerrung der Inferenz.



## **6 Diskussion**

### **6.1 Interpretation der Ergebnisse**

Tiefere Analyse der Ergebnisse und ihrer Implikationen.

### **6.2 Vergleich mit bestehenden Arbeiten**

Wie sich die Ergebnisse zu bereits veröffentlichten Forschungen verhalten.

### **6.3 Limitationen und Herausforderungen**

Kritische Betrachtung der Grenzen der Studie und mögliche Probleme.





## 7 Fazit

Zusammenfassung der wichtigsten Erkenntnisse Praktische Implikationen: Wie die Ergebnisse in der Praxis angewendet werden können. Empfehlungen für zukünftige Forschungen: Vorschläge für weiterführende oder ergänzende Studien.



## **8 Anhang**



## **9 Literaturverzeichnis**



# **A Appendix**

## **A.1 Supporting Data**

## **A.2 Some Code Listings**





## **Abbildungsverzeichnis**



## **Tabellenverzeichnis**



# Literatur

- [PB00] J. C. Pinheiro und D. Bates. *Mixed-Effects Models in S and S-PLUS*. New York: Springer, 2000. ISBN: 978-1-4419-0318-1.