

Konstruktion einfacher Konfidenzintervalle für Linearkombinationen von Erfolgswahrscheinlichkeiten

Diplomarbeit

vorgelegt von
Frank Konietschke
aus Seesen

angefertigt im
Institut für Mathematische Stochastik
der Georg-August-Universität Göttingen

2006

*Statistiken sind wie Zahlengebäude.
Sollen sie gut sein, brauchen sie - wie gute Häuser - ein solides Fundament,
klare Konturen und den Beweis, daß sie im Wandel der Zeiten ihren Wert behalten.
Es gibt aber auch schlechte Statistiken. Sie fallen zusammen wie Kartenhäuser.¹*

¹Paul Schnitker (*1927), dt. Unternehmer, 1973-88 Präs. Zentralverband d. Dt. Handwerks 285

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
1 Einleitung	1
1.1 Motivation	1
1.2 Aufbau der Arbeit	2
2 Vorstellung der Beispiele	3
2.1 Brustkrebs Diagnosestudie	3
3 Definitionen, Sätze und Notationen	5
3.1 Grundlagen	5
3.2 Notationen	6
3.3 Numerische Integration	7
3.4 Transformationsmethode	8
4 Konfidenzintervalle für π	11
4.1 Gütemaße	11
4.1.1 A-Priori-Maße	11
4.1.2 Exkurs: Entstehung der Zacken im Coverage-Prozess	15
4.1.3 A-Posteriori-Maße	17
4.1.4 Konfidenzintervalle mit $X=0$ (Null- bzw. Vollergebnis)	20
4.2 Konfidenzintervalle für Erfolgswahrscheinlichkeit π	21
4.3 Transformierte Statistiken	23
4.3.1 stetige Transformationen	23
4.4 Vergleich der Konfidenzintervalle	31
4.4.1 Coverage-Prozesse	31
4.4.2 Ergebnisse numerischer Integrationen	32
4.4.3 p -Konfidenz und p -Bias	36
4.5 Zusammenfassung und Motivation für Übertragung auf Linearkombinationen	38
5 Konfidenzintervalle für $p_1 - p_2$ (verbundene Stichproben)	39
5.1 Modell für zwei verbundene Stichproben	39
5.2 Gütemaße	40
5.3 Konstruktion der verbundenen Coverage-Kurven	45
5.4 Konfidenzintervalle	45
5.4.1 Programmierung des Score-Intervalls	47
5.4.2 Transformierte Konfidenzintervalle	48
5.5 Vergleich der Konfidenzintervalle	56
5.5.1 Coverage-Kurven	56
5.6 Numerische Ergebnisse	57

6	Konfidenzintervalle für $p_1 - p_2$ (unverbundene Stichproben)	61
6.1	Vereinfachtes Modell mit unabhängigen Stichproben bzw. Methoden	61
6.1.1	Gütemaße	61
6.2	Konfidenzintervalle für $p_\Delta = p_1 - p_2$	66
6.3	Transformierte Statistiken	67
6.4	Vergleich der Konfidenzintervalle	70
6.4.1	Coverage-Prozesse	71
7	Auswertung des Beispiels	73
8	Zusammenfassung und Ausblick	75
A	Grenzwertsätze	79
B	Software	81
B.1	Zweistichprobenproblem-verbundene Stichproben	81
	Literaturverzeichnis	84

Abbildungsverzeichnis

- 3.1 Darstellung der Trapezregel 7
- 3.2 Schematische Darstellung der Transformationsmethode 9

- 4.1 Grenzverlauf 15
- 4.2 (Anti)-Konservativität 16
- 4.3 Ergebnisse der c-Optimierung 31
- 4.4 Logit- und Probit- Transformationsfunktion 31
- 4.5 Coverage-Kurven für n=5 32
- 4.6 Coverage-Kurven für n=10 32
- 4.7 Längen-Kurven für n=5 33
- 4.8 Längen-Kurven für n=10 33
- 4.9 Ergebnisse numerischer Integrationen für 5 Stichprobenumfänge 35
- 4.10 p-Konfidenz für n=10 37
- 4.11 p-Konfidenz für n=50 37
- 4.12 p-Bias für n=10 37
- 4.13 p-Bias für n=50 37

- 5.1 Contour Plots des zweiten Iterationsschrittes mit $k=1/50$ 55
- 5.2 Ergebnisse der verbundenen k-Optimierung 55
- 5.3 Coverage-Kurven mit n=60 und Odds-Ratio $\theta = 1$ bei gegebenem p_2 56
- 5.4 Coverage-Kurven mit n=60 und Odds-Ratio $\theta = 3$ bei gegebenem p_2 56
- 5.5 Längenparaboloid und Coverage-Prozess Transformiert mit $\theta=1$ und n=40 58
- 5.6 Vergleich der verbundenen Verfahren 59

- 6.1 Anforderungen an die Stichprobenumfänge bei unverbundenen Stichproben 63
- 6.2 Ergebnismenge der unverbundenen Optimierung 70
- 6.3 Ergebnisse der unverbundenen k-Optimierung 70
- 6.4 Coverage-Kurven bedingt auf $p_2=0.1$ mit $n_1 = n_2 = 40$ 71
- 6.5 Coverage-Kurven bedingt auf $p_2=0.5$ mit $n_1 = n_2 = 40$ 71
- 6.6 Vergleich der unverbundenen Verfahren 72

- 8.1 Abkürzungen der untersuchten Statistiken 76

Tabellenverzeichnis

5.1	Numerische Integrationsergebnisse mit $n=25$ und gegebenem Odds-Ratio $\theta = 3$ und $p_2 = 0.5, 0.1$	57
8.1	Niveausimulationen mit 4 Methoden zum Vergleich von Erfolgswahrscheinlichkeiten und Compound-Symmetric-Kovarianzstruktur ($n=50, 100$)	77
8.2	Niveausimulationen mit 8 Methoden zum Vergleich der Erfolgswahrscheinlichkeiten und Compound-Symmetric-Kovarianzstruktur ($n=50, 100$)	77

1 Einleitung

1.1 Motivation

Die vorliegende Arbeit beschäftigt sich mit der Auswertung von Datensätzen, die dichotome Messungen (sogenannte $\{0, 1\}$ -Daten) enthalten. Dichotome Daten sind Klassierungsvariablen, die in nur zwei möglichen Ausprägungen auftreten können. Beispiele für diese sind Indikatoren für das Geschlecht oder Nebenwirkungen (JA/Nein). Aufgrund ihrer einfachen Struktur tauchen Binärvariablen in nahezu jeder klinischen Studie auf.

Eine vollständige Analyse eines Datensatzes verlangt neben den klassischen Testverfahren zur Überprüfung auf statistische Signifikanz auch die Angabe von Konfidenzintervallen. Ein Konfidenzintervall zeigt mehr Informationen auf als der sogenannte p -Wert, mit welchem Testentscheidungen getroffen werden können. Schlussfolgerungen aus der Interpretation von Konfidenzintervallen können verfälscht sein, wenn das verwendete Intervall keine guten Eigenschaften hat. Bekannte Größen für die Beschaffenheit von Konfidenzintervallen sind z.B. die Länge oder die Coverage-Probability¹ eines Konfidenzintervalls. Die Coverage-Probability eines Intervalls zum Niveau α sollte bei einer minimalen Länge möglichst $(1 - \alpha)$ betragen.

Eines der ersten berechneten Konfidenzintervalle für einen unbekannt Parameter ist das sogenannte Wald-Konfidenzintervall für eine Erfolgswahrscheinlichkeit (vgl. LAPLACE (1812) [17]), welches aus der Invertierung eines Wald-Tests entsteht. Der Vorteil eines Wald-Tests ist, dass die asymptotische Verteilung der Prüfgröße bekannt ist. Die Verteilung der Teststatistik konvergiert allerdings nur sehr langsam gegen die Grenzverteilung, was bedeutet, dass der Test für geringe Stichprobenumfänge nicht das Niveau einhält. Die Coverage-Probability der Wald-Konfidenzintervalle ist demnach sehr gering. Weitere Konfidenzintervalle, die mit einem deutlich höheren Programmieraufwand verbunden sind, sind invertierte Score-Tests. Der erste invertierte Score-Test eines verbundenen Zweistichprobenproblems wurde in der Arbeit von TANGO (1998) [26] veröffentlicht.

Im Einstichprobenproblem ist das Wilson-Intervall als invertierter Score-Test bekannt. Interpretiert man den Mittelpunkt eines Konfidenzintervalls als Punktschätzer für den unbekannt Parameter, dann zeigt sich, dass das Wilson-Intervall einen Schätzer für die wahre Erfolgswahrscheinlichkeit verwendet, der sich approximativ aus der gewöhnlichen relativen Häufigkeit und einer bestimmten imaginären Anzahl an zusätzlichen Erfolgen und Misserfolgen zusammensetzt.

AGRESTI und COULL (1998) [17] konstruierten aus diesem Schätzer ein neues Intervall, indem sie diesen Schätzer in die einfache Formel zur Berechnung des Wald-Konfidenzintervalls einsetzten und damit ein sehr gutes Intervall erhielten. Die Grundidee, die Schätzer durch imaginäre Beobachtungen zu verändern und diese in die Wald-Gleichungen einzusetzen, wurde auch auf verbundene und unverbundene Zweistichproben übertragen. Auch hier zeigen diese Intervalle sehr gute Eigenschaften. Die Agresti-Konfidenzintervalle sind nicht schwerer zu berechnen als die Wald-Intervalle und damit sogar in Grundvorlesungen über Statistik mit in den Lehrstoff integrierbar. Große Nachteile dieser Intervalle liegen darin, dass diese nicht bereichserhaltend sind und sich die Konstruktion nur im Einstichprobenproblem analytisch begründen lässt. Die Herleitung der

¹Überdeckungswahrscheinlichkeit

Agresti-Intervalle für zwei Stichproben erfolgt mit einem heuristischen Verfahren und lässt sich somit nicht auf mehrfaktorielle Anlagen übertragen. Ebenso gibt es für die invertierten Score-Tests kein einheitliches Muster um diese Intervalle zu berechnen. Die Berechnung folgt einem iterativem Algorithmus und ist somit ohne unterstützenden Computer kaum möglich.

Das Ziel dieser Arbeit soll sein, ein Verfahren zu konstruieren, mit welchem sich 'gute' und vor allem einfach zu berechnende Konfidenzintervalle für Linearkombinationen von Erfolgswahrscheinlichkeiten berechnen lassen.

1.2 Aufbau der Arbeit

Der Aufbau der Arbeit gliedert sich wie folgt:

Die Bewertung und Herleitung der Konfidenzintervalle zu einer und zwei verbundenen bzw. unverbundenen Stichproben wird in je einem Kapitel vorgenommen. Die Maße, anhand deren die Statistiken bewertet werden, werden dazu in jedem Kapitel vorgestellt und erläutert. Die neu definierten Maße zur Konservativität und Anti-Konservativität werden zusätzlich anhand einer Grafik genau erklärt. Anschließend werden die bisher bekannten Statistiken mit den neu hergeleiteten Verfahren verglichen.

Als weitere Motivation zur Theorie und Analyse von dichotomen Daten wird ein kleines Beispiel herangezogen und ausgewertet.

Weiterhin entstanden bei der Anfertigung der Arbeit zahlreiche Anwenderprogramme in R und in SAS-IML. Für alle in dieser Arbeit vorgestellten Verfahren sind daher Makros erstellt, die jeweils die Grenzen der Intervalle bzw. a-priori- oder a-posteriori-Maße für gegebene Beobachtungen berechnen. Exemplarisch sind Quellcodes in der Arbeit und im Anhang abgebildet.

2 Vorstellung der Beispiele

2.1 Brustkrebs Diagnosestudie

- Als Motivation zur Theorie dichotomer Daten soll ein kleines Beispiel diese Arbeit beginnen.

Es handelt sich dabei um eine Brustkrebs Diagnosestudie mit verbundenen Stichproben. In dem kleinen Ausschnitt aus dieser Studie soll es darum gehen, die Diagnose eines erfahrenen Radiologen mit einer computerunterstützten Diagnose (CAD) zu vergleichen. Das Ziel soll sein, die Sensitivitäten

$$\pi_i = P(X_i = \text{'Krank'} | D^+), \text{ wobei } i = \text{Radiologe oder CAD gilt,}$$

zu bewerten und zu vergleichen. Die Sensitivität bezeichnet die Wahrscheinlichkeit dafür, dass wirklich kranke Patientinnen auch als krank diagnostiziert werden.

Insgesamt nahmen 226 durch Biopsie gesichert kranke Patientinnen an dieser Studie teil.

- Die folgende 4-Felder-Tafel zeigt den Ausgang dieser Studie. Die Klassifizierung der Kranken erfolgte durch eine Eins, die der Gesunden durch eine Null ($1 \hat{=} \text{'krank'}$ und $0 \hat{=} \text{'gesund'}$).

		Radiologe (2)		Total
		0	1	
CAD (1)	0	33	44	77
	1	7	142	149
Total		40	186	226

- Verwendet man die gewöhnlichen relativen Häufigkeiten als Schätzer für die Sensitivität, dann ergeben sich folgende geschätzte Sensitivitäten:

$$\hat{\pi}_1 = \frac{149}{226} = 0.66 \quad \text{und} \quad \hat{\pi}_2 = \frac{186}{226} = 0.82.$$

- An diesem Punkt trifft man auf weitere Aufgaben:
 - Teste die Nullhypothese $H_0 : \pi_1 = \pi_2$ mit einem guten Test im Bezug auf Niveau und Power.
 - Berechne möglichst einfache Konfidenzintervalle mit guten Eigenschaften für π_1, π_2 und $\pi_1 - \pi_2$.

3 Definitionen, Sätze und Notationen

3.1 Grundlagen

Die Operatoren $E(\cdot)$, $Var(\cdot)$ seien der Erwartungswert bzw. die Varianz einer Verteilung. Die Kovarianz von Zufallsvariablen sei $Cov(\cdot, \cdot)$.

Die Abbildung $P: \Omega \rightarrow [0, 1]$ sei ein Wahrscheinlichkeitsmaß, welches die Axiome von Kolmogorow erfüllt. Die Erfolgswahrscheinlichkeit einer Bernoulliverteilung sei π .

- **Begriffe aus der Analysis**

Matrizen und Vektoren werden stets **fett** geschrieben. Eine Matrix $\mathbf{A} \in \mathbb{R}^{p \times q}$ hat dabei folgende Gestalt:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pq} \end{pmatrix} \in \mathbb{R}^{p \times q}.$$

Die Matrix $\mathbf{A}^t \in \mathbb{R}^{q \times p}$ heißt die zu \mathbf{A} transponierte Matrix.

Besondere Matrizen sind Projektoren. Diese sind idempotent und symmetrisch. Ein Projektor ist die Matrix \mathbf{P} , definiert durch

$$\mathbf{P}_a = \mathbf{I}_a - \frac{1}{a} \mathbf{J}_a \in \mathbb{R}^{a \times a}.$$

Die Matrizen \mathbf{I} bzw. \mathbf{J} kennzeichnen die Einheitsmatrix bzw. die 'J-Matrix', welche nur Einsen enthält.

Definition 3.1.1 (GRADIENT)

Die Abbildung $f: \mathbb{R}^d \rightarrow \mathbb{R}$ sei differenzierbar und es gelte $\mathbf{x} = (x_1, \dots, x_d)^t$. Der Vektor

$$f'(\mathbf{x}) = \frac{d}{d\mathbf{x}} f(\mathbf{x}) = \left(\frac{\partial}{\partial x_1} f(\mathbf{x}), \dots, \frac{\partial}{\partial x_d} f(\mathbf{x}) \right)^t$$

heißt Gradient von f .

Definition 3.1.2 (JACOBI-MATRIX)

Die Abbildung $\mathbf{g}: \mathbb{R}^d \rightarrow \mathbb{R}^k$ sei differenzierbar und es gelte $\mathbf{g} = (g_1, \dots, g_k)^t$. Die Matrix

$$\frac{d}{d\mathbf{x}} [\mathbf{g}(\mathbf{x})]^t = \begin{pmatrix} \frac{d}{d\mathbf{x}} g_1(\mathbf{x}) \\ \vdots \\ \frac{d}{d\mathbf{x}} g_k(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} g_1(\mathbf{x}) & \dots & \frac{\partial}{\partial x_d} g_1(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} g_k(\mathbf{x}) & \dots & \frac{\partial}{\partial x_d} g_k(\mathbf{x}) \end{pmatrix}$$

heißt Jacobi-Matrix von \mathbf{g} .

- **Begriffe aus der elementaren Statistik**

Definition 3.1.3 (KONFIDENZINTERVALL FÜR ERFOLGSWAHRSCHEINLICHKEIT)

Es seien $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{B}(\pi)$ und $X = \sum_{i=1}^n X_i \sim \text{BIN}(n, \pi)$. $\mathcal{U}(X)$ und $\mathcal{O}(X)$ seien zwei Statistiken. Das Intervall $\mathcal{KI}(X) = [\mathcal{U}(X), \mathcal{O}(X)]$ heißt zweiseitiges Konfidenzintervall mit Konfidenzniveau $1 - \alpha$, falls $\forall \alpha \in (0, 1)$ gilt:

$$P(\mathcal{U}(X) \leq \pi \leq \mathcal{O}(X)) = 1 - \alpha. \quad (3.1)$$

- **Begriffe aus der Mathematischen/ Asymptotischen Statistik**

Den Grundbaustein dieser Arbeit liefern Grenzwertsätze und Konvergenzen.

Definition 3.1.4 (VERTEILUNGSKONVERGENZ)

Eine Folge von Zufallsvektoren \mathbf{Y}_n konvergiert in Verteilung gegen einen Zufallsvektor \mathbf{Z} mit Verteilungsfunktion $F(\mathbf{z})$, falls

$$\lim_{n \rightarrow \infty} P(\mathbf{Y}_n \leq \mathbf{z}) = F(\mathbf{z})$$

für jeden Stetigkeitspunkt \mathbf{z} gilt.

Schreibweise:

$$\mathbf{Y}_n \xrightarrow{\mathcal{L}} \mathbf{Z} \sim F(\mathbf{z}), \quad (n \rightarrow \infty).$$

Definition 3.1.5 (KONVERGENZ IN WAHRSCHEINLICHKEIT)

\mathbf{Y}_n konvergiert in Wahrscheinlichkeit gegen \mathbf{Z} , $\mathbf{Y}_n \xrightarrow{P} \mathbf{Z}$, falls $\forall \varepsilon > 0$ gilt:

$$P(|\mathbf{Y}_n - \mathbf{Z}| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

3.2 Notationen

In der Arbeit werden primär bedingte Erwartungswerte berechnet. Dieses bedeutet, dass Erwartungswerte von Zufallsvariablen unter der Voraussetzung berechnet werden, dass bestimmte Parameter bekannt sind. Gekennzeichnet werden diese Parameter als Indizes. Im Verlauf der Arbeit wird, um Unverständnissen vorzubeugen, häufiger auf diese Gegebenheit hingewiesen. Mißverständnisse können durch diese Notation eigentlich nicht entstehen.

3.3 Numerische Integration

Fast alle in dieser Arbeit gewonnenen Resultate sind Ergebnisse numerischer Integrationen. Numerische Integration findet da Anwendung, wo Funktionen nicht elementar integrierbar sind. Die Grundidee dabei ist, den Integranden hinreichend genau durch Funktionen zu approximieren, die sich leicht elementar integrieren lassen.

Verschiedene Ansätze zur Theorie der numerischen Integration sind in nahezu jedem Standard- bzw. Nachschlagewerk über höhere Mathematik zu finden.

Die einfachste Methode ist die Approximation des Integranden durch einen Streckenzug. Die Abbildung $f : [a, b] \rightarrow \mathbb{R}$ sei stetig und $n \in \mathbb{N} - \{0\}$. Das Intervall $[a, b]$ wird in n gleich große Teile durch die Teilpunkte

$$x_k := a + kh, \text{ wobei } h = (b - a)/n \text{ und } k = 0, \dots, n \text{ gilt,}$$

geteilt. Im Intervall $[x_k, x_{k+1}]$ wird f durch die lineare Funktion

$$x \mapsto f(x_k) + \frac{1}{h}(f(x_{k+1}) - f(x_k))(x - x_k), \text{ für } k = 0, \dots, n - 1,$$

approximiert, die an den Endpunkten x_k und x_{k+1} dieselben Werte wie f hat.

Die Funktionenfolge $g_n : [a, b] \rightarrow \mathbb{R}$ sei weiterhin der gesamte Streckenzug.

Da f gleichmäßig stetig ist, konvergiert die Folge g_n gleichmäßig gegen f und damit folgt:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} T_n,$$

wobei

$$\begin{aligned} T_n &:= T_n(f; a, b) := \int_a^b g_n(x) dx = \sum_{k=0}^{n-1} \frac{1}{2} (f(x_k) + f(x_{k+1})) \\ &= h \left(\frac{1}{2} f(x_0) + f(x_1) + \dots + f(x_{n-1}) + \frac{1}{2} f(x_n) \right) \end{aligned}$$

gilt.

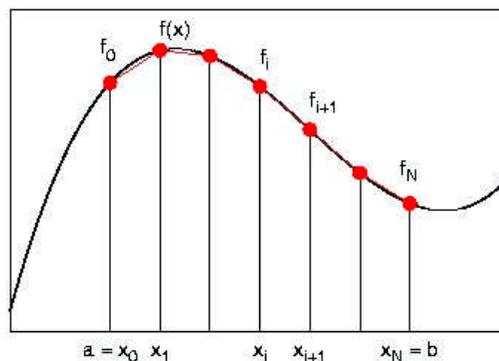


Abbildung 3.1: Darstellung der Trapezregel

Bivariate numerische Integrationen, also Ausdrücke der Form

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_1 dx_2,$$

lassen sich leicht numerisch approximieren, indem die Trapezregel für beide Integrale angewendet wird. Man sagt dann, dass die zusammengesetzte Trapezregel verwendet wird.

Vorbemerkt sei hier, dass in dieser Arbeit unstetige Funktionen numerisch integriert werden. Die Anwendung der obigen Verfahren ist gerechtfertigt, da die Funktionen vollständig auf $[0,1]$ beschränkt sind (d.h., es gibt kein Polverhalten im Definitionsbereich). Damit sind notwendigerweise auch die Sprunghöhen auf $[0,1]$ beschränkt. Bei einer möglichst hohen Anzahl an Integrationsstützstellen ist damit das vorgestellte Verfahren ein guter Schätzer für

$$\int_a^b f(x) dx.$$

3.4 Transformationsmethode

Unter der Transformationsmethode versteht man ein asymptotisches Verfahren, anhand dessen mit Hilfe stetiger Transformationen und Anwendung der δ -Methode (siehe Theorem A.0.8) Statistiken modifiziert werden können. Ausführlich wurde dieses Thema in den Arbeiten von DOMHOF (2001) [13] und CHRISTOPHLIEMK (2001) [12] behandelt.

In dieser Arbeit werden Konfidenzintervalle für verschiedene Linearkombinationen von Erfolgswahrscheinlichkeiten mit Hilfe der Transformationsmethode hergeleitet.

Definition 3.4.1 (*LINEARKOMBINATION*)

Es gelte $\mathbf{p} = (p_1, p_2, \dots, p_d)^t \in (0, 1)^d$ und $a_i \in \mathbb{R}$ für $i = 1, \dots, d$. Die Abbildung

$$\mathcal{L} : (0, 1)^d \rightarrow \mathbb{R}, \quad \mathbf{p} \mapsto \sum_{i=1}^d a_i p_i,$$

heißt *Linearkombination der Erfolgswahrscheinlichkeiten* p_i . Der Wertebereich von \mathcal{L} heißt Θ und wird auch *Parameterraum* von \mathcal{L} genannt.

$\widehat{\mathcal{L}}$ bezeichnet außerdem einen erwartungstreuen Schätzer für \mathcal{L} .

Das Prinzip der Transformationsmethode liegt darin, mit Hilfe einer umkehrbaren Funktion $g : \Theta \rightarrow \mathbb{R}$ die offene Menge Θ auf die Menge der reellen Zahlen abzubilden. Unter bestimmten Voraussetzungen an die Transformation $g : \Theta \rightarrow \mathbb{R}$ besitzt die Statistik

$$\sqrt{n}[g(\widehat{\mathcal{L}}) - g(\mathcal{L})]$$

eine asymptotische Normalverteilung, so dass ein asymptotisches Konfidenzintervall $[\mathcal{U}^{Transf}, \mathcal{O}^{Transf}]$ für $g(\mathcal{L})$ angegeben werden kann. Wegen der Umkehrbarkeit von g folgt, dass $g^{-1}([\mathcal{U}^{Transf}, \mathcal{O}^{Transf}])$ ein asymptotischer Konfidenzbereich für \mathcal{L} ist.

Damit $g^{-1}([\mathcal{U}^{Transf}, \mathcal{O}^{Transf}])$ wieder ein Intervall darstellt, sollte die Funktion g streng monoton sein. Die nächste Auflistung zeigt vier Anforderungen, die an eine geeignete Transformationsfunktion $g : \Theta \rightarrow \mathbb{R}$ vorerst gestellt werden müssen:

(T1) g sei stetig,

(T2) g sei bijektiv,

(T3) $g'(\pi) \neq 0 \forall \pi \in \Theta$,

(T4) g sei streng monoton steigend.

Für die Transformation g kommen damit alle streng monotonen und stetig differenzierbaren Bijektionen $g : \Theta \rightarrow \mathbb{R}$ in Frage, deren erste Ableitungen nullstellenfrei sind. In der Arbeit wird häufig auf die Anforderungen (T1)-(T4) verwiesen. Aus vereinfachenden Gründen werden die Anforderungen (T1)-(T4) unter dem Sammelbegriff 'Anforderungen an eine Transformation' zusammengefasst.

Eine weitere Einschränkung der Transformationsfunktionen erfolgt in dem jeweiligen Kapitel. Das einfache Prinzip der Transformationsmethode für Linearkombinationen zeigt Abbildung 3.2.

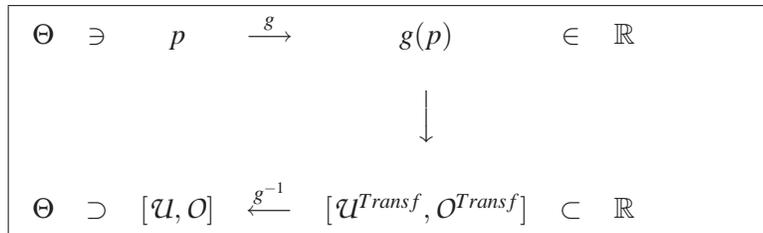


Abbildung 3.2: Schematische Darstellung der Transformationsmethode

4 Konfidenzintervalle für π

Als Grundlage zur Analyse von Konfidenzintervallen für eine Erfolgswahrscheinlichkeit dient ein Bernoullimodell:

$$X_1, X_2, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(\pi) \Rightarrow X = \sum_{j=1}^n X_j \sim \text{BIN}(n, \pi). \quad (4.1)$$

4.1 Gütemaße

4.1.1 A-Priori-Maße

Im Einstichprobenproblem wird zwischen a-priori- und a-posteriori-Maßen unterschieden. A-priori-Maße sind stets hypothetische Größen, die eine Statistik vor Beobachtung bewerten.

Definition 4.1.1 (LÄNGENKURVE)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. $\mathcal{KI}(X) = [\mathcal{U}(X), O(X)]$ sei ein Konfidenzintervall für π . Die Abbildung

$$L_n : (0, 1) \rightarrow [0, 1], \quad \pi \mapsto E(O(X) - \mathcal{U}(X)),$$

heißt Längenkurve.

Lemma 4.1.2 (ERWARTETE LÄNGE)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. Die Abbildung $L_n : (0, 1) \rightarrow [0, 1]$ sei eine Längenkurve. Für $\pi \in (0, 1)$ gilt dann:

$$L_n(\pi) = \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \cdot \binom{n}{k} \pi^k (1 - \pi)^{n-k}. \quad (4.2)$$

Beweis:

Sei X binomialverteilt zu n und π , dann gilt:

$$\begin{aligned} E(O(X) - \mathcal{U}(X)) &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \cdot P(X = k) \\ &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \cdot \binom{n}{k} \pi^k (1 - \pi)^{n-k}. \end{aligned}$$

□

Lemma 4.1.3 (EIGENSCHAFTEN DER LÄNGENKURVE)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. Die Abbildung $L_n : (0, 1) \rightarrow [0, 1]$ sei eine Längenkurve des Konfidenzintervalls aus Definition 4.1.1. Dann gilt:

- (a) L_n ist zweimal differenzierbar auf $(0, 1)$ und damit auch stetig,
 (b) L_n ist symmetrisch um $p = \frac{1}{2}$,
 (c) L_n beschreibt eine nach unten geöffnete Parabel mit Scheitel in $A = (\frac{1}{2}, L_n(\frac{1}{2}))$.

Beweis:

- (a) Die Differenzierbarkeit ergibt sich aus der Differenzierbarkeit der Summanden.
 (b) Es gelte $\pi = \frac{1}{2}$ und sei $\delta \in (0, \frac{1}{2})$ gegeben, dann gilt:

$$\begin{aligned} L_n\left(\frac{1}{2} + \delta\right) &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \cdot \binom{n}{k} \left(\frac{1}{2} + \delta\right)^k \left(\frac{1}{2} - \delta\right)^{n-k} \\ &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \cdot \binom{n}{k} \left(\frac{1}{2} + \delta\right)^{n-k} \left(\frac{1}{2} - \delta\right)^k \\ &= L_n\left(\frac{1}{2} - \delta\right). \end{aligned}$$

- (c) Notwendige und hinreichende Bedingung für ein Maximum:

$$\frac{dL_n}{d\pi}(\pi^*) = 0 \quad \text{und} \quad \frac{d^2L_n}{d^2\pi}(\pi^*) < 0. \quad (4.3)$$

Da eine Ableitung ein linearer Operator ist, ist die Ableitung einer Summe die Summe der Ableitungen. Demnach gilt:

$$\begin{aligned} \frac{dL_n(\pi)}{d\pi} &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \frac{d}{d\pi} \left(\pi^k (1 - \pi)^{n-k} \right) \\ &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \pi^{k-1} (1 - \pi)^{n-k-1} (k(1 - \pi) - \pi(n - k)), \\ \frac{d^2L_n(\pi)}{d^2\pi} &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \frac{d}{d\pi} \left(\pi^{k-1} (1 - \pi)^{n-k-1} (k(1 - \pi) - \pi(n - k)) \right) \\ &= \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \pi^{k-2} (1 - \pi)^{n-k-2} \times \\ &\quad \left(k(k-1)(1 - \pi)^2 - 2k(n-k)\pi(1 - \pi) + \pi^2(n-k)(n-k-1) \right). \end{aligned}$$

Mit diesen Ableitungen müssen die notwendige und die hinreichende Bedingung (4.3) für das Vorliegen eines Extremums überprüft werden. Es folgt weiterhin:

$$\begin{aligned} &\frac{dL_n(\pi)}{d\pi} = 0 \\ \iff &\sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \pi^{k-1} (1 - \pi)^{n-k-1} (k(1 - \pi) - \pi(n - k)) = 0 \\ \iff &\sum_{k=0}^n \pi^{k-1} (1 - \pi)^{n-k-1} (k(1 - \pi) - \pi(n - k)) = 0 \\ \iff &\pi^* = \frac{1}{2}. \end{aligned}$$

Speziell an der Stelle $\pi^* = \frac{1}{2}$ erhält man für die zweite Ableitung:

$$\frac{d^2 L_n}{d^2 \pi} \left(\pi^* = \frac{1}{2} \right) = \left(\frac{1}{2} \right)^{n-2} \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} [(n-2k)^2 - n].$$

Da $[O(k) - \mathcal{U}(k)] < 1 \forall k \in \mathbb{N}$ gilt, läßt sich weiterhin abschätzen:

$$\begin{aligned} & \left(\frac{1}{2} \right)^{n-2} \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} [(n-2k)^2 - n] \\ < & \left(\frac{1}{2} \right)^{n-2} \sum_{k=0}^n \binom{n}{k} ((n-2k)^2 - n) \\ = & \left(\frac{1}{2} \right)^{n-2} \sum_{k=0}^n \binom{n}{k} [(n-2k)^2] - \left(\frac{1}{2} \right)^{n-2} n \sum_{k=0}^n \binom{n}{k} \\ = & \left(\frac{1}{2} \right)^{n-2} [n^2 \cdot 2^n - 4n^2 \cdot 2^{n-1} + 4n(n+1) \cdot 2^{n-2}] - 4n \\ = & 4n - 4n \\ = & 0. \end{aligned}$$

Man erhält damit, wie behauptet, ein Maximum an der Stelle:

$$A = \left(\frac{1}{2}, L_n \left(\frac{1}{2} \right) \right) = \left(\frac{1}{2}, \sum_{k=0}^n [O(k) - \mathcal{U}(k)] \binom{n}{k} \cdot \left(\frac{1}{2} \right)^n \right).$$

□

Definition 4.1.4 (MITTLERE LÄNGE)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. $\mathcal{KI}(X)$ sei ein Konfidenzintervall für π . Die Abbildung $L_n : (0, 1) \rightarrow [0, 1]$ sei eine Längenkurve. Die Größe

$$\bar{L}_n = \int_0^1 L_n(\pi) d\pi \tag{4.4}$$

heißt mittlere Länge des Konfidenzintervalls \mathcal{KI} .

Definition 4.1.5 (COVERAGE-PROBABILITY)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. $\mathcal{KI}(X)$ sei ein Konfidenzintervall für π . Das Maß

$$P(\pi \in \mathcal{KI}(X))$$

heißt Coverage-Probability des Konfidenzintervalls \mathcal{KI} für festes π . Die Abbildung

$$C_n : (0, 1) \rightarrow [0, 1], \quad \pi \mapsto P(\pi \in \mathcal{KI}(X)), \tag{4.5}$$

heißt Coverage-Kurve bzw. Coverage-Prozess.

Lemma 4.1.6 (EXAKTE COVERAGE-PROBABILITY)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. Die Abbildung $C_n : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Konfidenzintervalls \mathcal{KI} . Für $\pi \in (0, 1)$ gilt dann:

$$C_n(\pi) = P(\pi \in \mathcal{KI}(X)) = \sum_{k=0}^n \mathbb{I}_{\pi \in \mathcal{KI}(k)} \binom{n}{k} \pi^k (1 - \pi)^{n-k}. \quad (4.6)$$

Die Abbildung $\mathbb{I}_{\pi \in \mathcal{KI}(X)}$ bezeichnet eine Indikatorabbildung, mit

$$\mathbb{I}_{\pi \in \mathcal{KI}(X)} = \begin{cases} 1, & \text{falls } \pi \in \mathcal{KI}(X), \\ 0, & \text{falls } \pi \notin \mathcal{KI}(X). \end{cases}$$

Beweis:

Es gilt:

$$P(\pi \in \mathcal{KI}(X)) = E(\mathbb{I}_{\pi \in \mathcal{KI}(X)}) = \sum_{k=0}^n \mathbb{I}_{\pi \in \mathcal{KI}(k)} \cdot P(X = k) = \sum_{k=0}^n \mathbb{I}_{\pi \in \mathcal{KI}(k)} \binom{n}{k} \pi^k (1 - \pi)^{n-k}.$$

□

Lemma 4.1.7 (EIGENSCHAFTEN DES COVERAGE-PROZESSES)

Es gelte das Modell (4.1). Die Abbildung $C_n : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Konfidenzintervalls aus Definition 4.1.5. Dann gilt:

- (1) Die Funktion $C_n : (0, 1) \rightarrow [0, 1]$, $\pi \mapsto P(\pi \in \mathcal{KI}(X))$, ist wohldefiniert.
- (2) C_n ist unstetig.
- (3) C_n ist symmetrisch um $\pi = \frac{1}{2}$.

Beweis:

$$(1) \quad 0 \leq C_n(\pi) = \sum_{k=0}^n \mathbb{I}_{\pi \in \mathcal{KI}(k)} \binom{n}{k} \pi^k (1 - \pi)^{n-k} \leq \sum_{k=0}^n \binom{n}{k} \pi^k (1 - \pi)^{n-k} = 1.$$

(2) Der Indikator ist unstetig. Damit ist auch C_n unstetig.

(3) Es gelte $\pi = \frac{1}{2}$ und sei $\delta \in (0, \frac{1}{2})$ gegeben, dann gilt:

$$\begin{aligned} C_n\left(\frac{1}{2} + \delta\right) &= \sum_{k=0}^n \mathbb{I}_{\frac{1}{2} + \delta \in \mathcal{KI}(k)} \binom{n}{k} \left(\frac{1}{2} + \delta\right)^k \left(1 - \left(\frac{1}{2} + \delta\right)\right)^{n-k} \\ &\stackrel{(4.1.3)}{=} \sum_{k=0}^n \mathbb{I}_{\frac{1}{2} - \delta \in \mathcal{KI}(k)} \binom{n}{k} \left(\frac{1}{2} - \delta\right)^k \left(\frac{1}{2} + \delta\right)^{n-k} \\ &= C_n\left(\frac{1}{2} - \delta\right). \end{aligned}$$

□

Damit wurde eine a-priori Größe geschaffen, mit welcher die Qualität eines Konfidenzintervalls genau bewertet werden kann.

Aus der diskreten Struktur des unterlegten Modells folgt, dass die Coverage-Kurve eines Konfidenzintervalls \mathcal{KI} zum Niveau α nicht konstant $(1 - \alpha)$ sein kann. Allerdings sollte das Intervall, um Definition 3.1.3 zu erfüllen, eine Coverage-Kurve erzeugen, die nicht stark von $(1 - \alpha)$ abweicht. Das Bild eines Coverage-Prozesses zeigt eine Zackenkurve über $\Theta = \{0 < \pi < 1\}$.

4.1.2 Exkurs: Entstehung der Zacken im Coverage-Prozess

Die Entstehung der Zacken läßt sich analytisch folgendermaßen erklären:

Als Beispiel dazu zeigt die Abbildung 4.1 das Bild vom Verlauf der Ober- bzw. Untergrenze eines beliebigen Konfidenzintervalls. Es gelte $\pi_1 = 1/2$ und $\mathcal{K}I(k)$, $k \in \mathbb{N}$, sei ein aus k konstruiertes Konfidenzintervall für π .

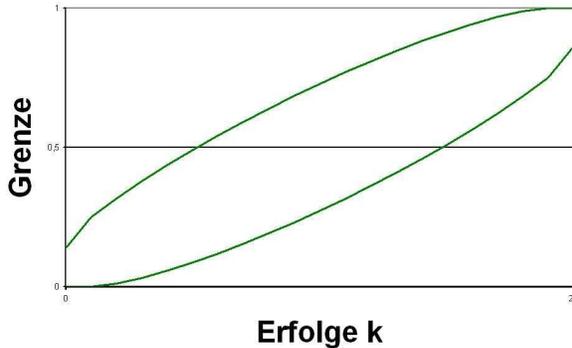


Abbildung 4.1: Grenzverlauf

Gesucht sei $C_n(1/2)$. Für $k = 0$ ist π_1 nicht in $\mathcal{K}I(0)$, weil hier die Obergrenze über π_1 liegt. Damit nimmt hier der Indikator den Wert Null an. Ab einem Wert von $O(k) > \frac{1}{2}$ liegt die Obergrenze stets über π_1 . Zu betrachten ist weiterhin die Untergrenze. Der Indikator nimmt genau dann den Wert Eins an, falls $O(k) - \pi_1 \geq 0$ und $\pi_1 - \mathcal{U}(k) \geq 0$ gilt. Damit wird der Indikatorfunktion der Wert Null zugeordnet, sobald die Untergrenze größer als $\pi_1 = 1/2$ ist. Da die Ober- bzw. Untergrenzen mit wachsendem k nicht konstant verlaufen, greift der Indikator an stets unterschiedlichen Stellen, wodurch die charakteristischen Zacken entstehen.

Um das Gesamtverhalten des Coverage-Prozesses in einem Ausdruck zu charakterisieren, hilft folgende Definition:

Definition 4.1.8 (MITTLERE COVERAGE-PROBABILITY)

Es gelte das Modell (4.1). Die Abbildung $g : [0, 1] \rightarrow \mathbb{R}^+$ sei eine Wahrscheinlichkeitsdichte und $C_n : (0, 1) \rightarrow [0, 1]$ ein Coverage-Prozess des Konfidenzintervalls $\mathcal{K}I$. Die Größe

$$\bar{C}_n = \int_0^1 C_n(\pi)g(\pi)d\pi \quad (4.7)$$

heißt *mittlere Coverage-Probability des Intervalls $\mathcal{K}I$* .

Die Definition 4.1.8 stellt damit die erwartete Überdeckungswahrscheinlichkeit eines Konfidenzintervalls $\mathcal{K}I$ dar. Am sinnvollsten ist es, alle Werte aus dem $(0, 1)$ Intervall als gleichwahrscheinlich anzusehen. Unter dieser Annahme wird $g(\pi)$ zur Dichtefunktion der stetigen Gleichverteilung auf $[0, 1]$. Die Dichtefunktion dieser Verteilung ist konstant Eins.

In all den folgenden Definitionen wird der Funktion $g(\pi)$ diese Verteilung zugeschrieben und kann deshalb auch aus vereinfachenden Gründen weggelassen werden.

Anschaulich bedeutet Definition 4.1.8, dass die mittlere Coverage-Probability als die Gesamtfläche zwischen der Abszisse und der Coverage-Kurve angesehen wird.

Auf diesen Ideen aufbauend lassen sich neue Maße konstruieren und definieren.

Betrachtet man einen Coverage-Prozess, dann zeigt dieser das Verhalten des zugrundegelegten Konfidenzintervalls auf. Weiterhin sagt man, dass dieser bei Unterschreitung des Nominalniveaus α konservatives bzw. anti-konservatives Verhalten bei Überschreitung des Nominalniveaus aufweist. Diese Formulierung ist äquivalent zu der Berechnung der Flächen zwischen Nominallinie $1 - \alpha$ und der Coverage-Kurve und damit zur Konstruktion von Maßen, die die Konservativität bzw. Anti-Konservativität a-priori messen.

Definition 4.1.9 (MITTLERE KONSERVATIVITÄT und ANTI-KONSERVATIVITÄT)

Es gelte das Modell (4.1). Die Abbildung $g : [0, 1] \rightarrow \mathbb{R}^+$ sei eine Wahrscheinlichkeitsdichte und $C_n : (0, 1) \rightarrow [0, 1]$ ein Coverage-Prozess des Konfidenzintervalls $\mathcal{K}I$. Die Größen

$$\bar{K}_n = \int_{\{C_n(\pi) \geq 1-\alpha\}} C_n(\pi)g(\pi)d\pi \quad \text{bzw.} \quad (4.8)$$

$$\overline{AK}_n = \int_{\{C_n(\pi) \leq 1-\alpha\}} C_n(\pi)g(\pi)d\pi \quad (4.9)$$

heißen mittlere Konservativität bzw. mittlere Anti-Konservativität des Intervalls $\mathcal{K}I$.

Konservativität Anti-Konservativität

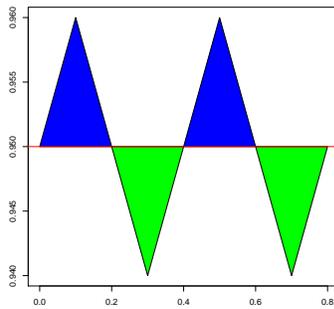


Abbildung 4.2: (Anti-)Konservativität

Abbildung 4.2 zeigt eine stark vereinfachte Abbildung einer Coverage-Kurve eines beliebigen Konfidenzintervalls $\mathcal{K}I$ zum Niveau $\alpha = 0.05$. Deutlich sind die charakteristischen Zacken zu sehen. Nach Definition 4.1.9 entspricht die Konservativität der Gesamtfläche oberhalb der Coverage-Kurve, die Anti-Konservativität der Gesamtfläche unterhalb des Coverage-Prozesses. In Abbildung 4.2 sind diese Größen blau bzw. grün dargestellt. Mit diesen Maßen wird ein Konfidenzintervall als 'gut' bewertet, wenn sowohl die Konservativität, als auch die Anti-Konservativität sehr gering sind. Weiterhin werden Größen berechnet, die den Abstand der exakten Coverage-Probability, sowie deren Streuung um das Konfidenzniveau $1 - \alpha$ messen.

Definition 4.1.10 (MITTLERER ABSOLUTER FEHLER)

Es gelte das Modell (4.1). Die Abbildung $C_n : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Intervalls $\mathcal{K}I$ für π . Die Größe

$$\bar{F}_n = \int_0^1 |C_n(\pi) - (1 - \alpha)|d\pi \quad (4.10)$$

heißt mittlerer Fehler des Intervalls $\mathcal{K}I$.

Der mittlere Fehler beschreibt den relativen Abstand des Coverage-Prozesses zum $(1 - \alpha)$ Konfidenzniveau.

Definition 4.1.11 (STREUUNG)

Es gelte das Modell (4.1). Die Abbildung $C_n : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Intervalls $\mathcal{K}I$. Die Größe

$$\bar{S}_n = \sqrt{\int_0^1 [C_n(\pi) - (1 - \alpha)]^2 d\pi} \quad (4.11)$$

heißt Streuung des Coverage-Prozesses um $1 - \alpha$.

Die Streuung beschreibt demnach die Quadratwurzel des mittleren quadratischen Fehlers¹ des Konfidenzintervalls $\mathcal{K}I$.

¹Allgemein ist dieser definiert durch $E(\hat{v} - v)^2$ eines unbekanntem Parameters v und zugehörigem Schätzer \hat{v} .

Definition 4.1.12 (ABWEICHUNG)

Es gelte das Modell (4.1). Die Abbildung $C_n : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Intervalls \mathcal{KI} und \bar{C}_n die mittlere Coverage-Probability. Die Größe

$$\bar{V}_n = \sqrt{\int_0^1 [C_n(\pi) - \bar{C}_n]^2 d\pi} \quad (4.12)$$

heißt Abweichung des Coverage-Prozesses.

4.1.3 A-Posteriori-Maße

Analysiert und vorgestellt werden nun Größen zur Bewertung von Statistiken nach Beobachtung des Erfolgs k einer binomialverteilten Zufallsvariable zu n und π . Als Grundlage dazu dient ein beliebiger Hypothesentest, der folgende Nullhypothese testet:

$$H_0 : \pi = \pi_0. \quad (4.13)$$

In der Arbeit von VOS und HUDSON (2005) [15] wird die Konstruktion von a-posteriori-Maßen durch die Interpretation eines Konfidenzintervalls als die Menge aller Werte, die von einem zum Konfidenzintervall \mathcal{KI} korrespondierenden Hypothesentest $H_0 : \pi = \pi_0$ nicht abgelehnt werden, begründet.

[...] It is commonly accepted that confidence sets and tests are two sides of the same coin: the confidence set consists of those parameter values that are not rejected by a test at the corresponding level (Skovgaard 2001, p.4).

Testentscheidungen werden in der Regel mit Hilfe des sogenannten p-Wertes getroffen. Der p-Wert gibt die Wahrscheinlichkeit an, dass (unter Annahme der Gültigkeit der Nullhypothese H_0) die Teststatistik den beobachteten oder einen extremeren Wert annimmt.

Definition 4.1.13 (2-SEITIGE p-WERT FUNKTION)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. $P_\pi(X \leq k)$ sei die Verteilungsfunktion von X . Die Abbildung

$$A_k : (0, 1) \rightarrow [0, 1], \quad \pi \mapsto A_k(\pi) = \min\{2 \cdot P_\pi(X \leq k), 2 \cdot P_\pi(X \geq k), 1\},$$

heißt 2-seitige p-Wert Funktion. Insbesondere gilt für $\pi \in (0, 1)$:

$$A_k(\pi) = \min \left\{ 2 \cdot \sum_{i=0}^k \binom{n}{i} \cdot \pi^i \cdot (1-\pi)^{n-i}, 2 \cdot \left(1 - \sum_{i=0}^{k-1} \binom{n}{i} \cdot \pi^i \cdot (1-\pi)^{n-i} \right), 1 \right\}.$$

Beachte: Die 2-seitige p-Wert Funktion ist eine Funktion von π , abhängig von k .

Lemma 4.1.14 (EIGENSCHAFTEN DER 2-SEITIGEN p-WERT FUNKTION)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. $P_\pi(X \leq k)$ sei die Verteilungsfunktion von X und $A_k : (0, 1) \rightarrow [0, 1]$ die 2-seitige p-Wert Funktion. Dann gilt:

- (1) $A_k(\pi)$ ist stetig $\forall k \in \mathbb{N}$,
- (2) $A_k(\pi)$ ist monoton steigend für $\pi < \frac{k}{n}$,
- (3) $A_k(\pi)$ ist monoton fallend für $\pi > \frac{k}{n}$,

$$(4) A_k(\pi = \frac{k}{n}) = 1 \forall k \in \mathbb{N}.$$

Beweis:

(1) Die Stetigkeit folgt aus der Stetigkeit des Absolutbetrages, wobei

$$\begin{aligned} A_k(\pi) &= \min\{2 \cdot P_\pi(X \leq k), 2 \cdot P_\pi(X \geq k), 1\} \\ &= \min\{\underbrace{\min\{2 \cdot P_\pi(X \leq k), 2 \cdot P_\pi(X \geq k)\}}_{=:J}, 1\} \\ &= \frac{1}{2} \cdot [J + 1 - |J - 1|] \end{aligned}$$

gilt. Für J gilt insbesondere:

$$J = \frac{1}{2} \cdot [2P_\pi(X \leq k) + 2P_\pi(X \geq k) - |2P_\pi(X \leq k) - 2P_\pi(X \geq k)|].$$

(2) Es gelte $0 < \pi_1 < \pi_2 < \frac{k}{n}$. Dann gilt:

$$A_k(\pi_1) = \min\{2P_{\pi_1}(X \leq k), 2P_{\pi_1}(X \geq k), 1\}.$$

Aufgrund der Binomialverteilung und mit Hilfe der Darstellung von (1) gilt:

$$\begin{aligned} A_k(\pi_1) &= \min\{2P_{\pi_1}(X \leq k), 2P_{\pi_1}(X \geq k), 1\} \\ &= \min\left\{2 \cdot \sum_{i=0}^k \binom{n}{i} \cdot \pi_1^i \cdot (1 - \pi_1)^{n-i}, 2 \cdot \left(1 - \sum_{i=0}^{k-1} \binom{n}{i} \cdot \pi_1^i \cdot (1 - \pi_1)^{n-i}\right), 1\right\} \\ &\leq \min\left\{2 \cdot \sum_{i=0}^k \binom{n}{i} \cdot \pi_2^i \cdot (1 - \pi_2)^{n-i}, 2 \cdot \left(1 - \sum_{i=0}^{k-1} \binom{n}{i} \cdot \pi_2^i \cdot (1 - \pi_2)^{n-i}\right), 1\right\} \\ &= A_k(\pi_2). \end{aligned}$$

(3) Der Beweis verläuft analog zu (2).

(4) Es gilt:

$$A_k\left(\pi = \frac{k}{n}\right) = \min\left\{2P_{\frac{k}{n}}(X \leq k), 2P_{\frac{k}{n}}(X \geq k), 1\right\}.$$

Weiterhin folgt:

$$\begin{aligned} 2P_{\frac{k}{n}}(X \leq k) &= 2 \cdot \sum_{j=0}^k \binom{n}{j} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} = \frac{2}{n^n} \underbrace{\sum_{j=0}^k \binom{n}{j} k^j (n-k)^{n-j}}_{\geq \frac{1}{2}n^n} \geq 1, \\ 2P_{\frac{k}{n}}(X \geq k) &= 2 \cdot \sum_{j=k}^n \binom{n}{j} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} = \frac{2}{n^n} \underbrace{\sum_{j=k}^n \binom{n}{j} k^j (n-k)^{n-j}}_{\geq \frac{1}{2}n^n} \geq 1. \end{aligned}$$

□

Definition 4.1.15 (*p-KONFIDENZ*)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$, wobei die Beobachtung von X k sei. $A_k(\pi)$ sei die 2-seitige p -Wert Funktion und $[\mathcal{KI}(k)] = [\mathcal{U}(k), O(k)]$ ein Konfidenzintervall für π . Die Zahl

$$c_n(\mathcal{KI}(k)) = \left(1 - \sup_{\pi \notin \mathcal{KI}(k)} A_k(\pi) \right)$$

heißt p -Konfidenz des Intervalls $\mathcal{KI}(k)$.

Bekanntlich schließt ein Konfidenzintervall für einen Parameter π einen Bereich ein, der mit einer zuvor festgelegten Wahrscheinlichkeit den wahren Parameter enthält. Testet man nun die Nullhypothese $H_0 : \pi = \pi_0$, dann ist das Kriterium für das Ablehnen der Nullhypothese bei Signifikanzniveau α , ob das entsprechende $(1 - \alpha)$ -Konfidenzintervall den Wert π_0 enthält oder nicht. Die p -Konfidenz beschreibt, wie stark durch die Beobachtung k das Intervall $[\mathcal{KI}(k)] = [\mathcal{U}(k), O(k)]$ Werte π_0 ablehnt, die nicht im Intervall enthalten sind. Zu einem gegebenen Niveau α sollte demnach die p -Konfidenz des Konfidenzintervalls $\mathcal{KI}(k)$ genau $(1 - \alpha)$ betragen.

Satz 4.1.16 (*BERECHNUNG DER p-KONFIDENZ*)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$, wobei die Beobachtung von X k sei. $P_\pi(X \leq k)$ sei die Verteilungsfunktion von X und die Größe $c_n(\mathcal{KI}(k))$ die p -Konfidenz des Konfidenzintervalls $\mathcal{KI}(k) = [\mathcal{U}(k), O(k)]$. $A_k(\pi)$ sei die 2-seitige p -Wert Funktion. Dann gilt:

$$\begin{aligned} c_n(\mathcal{KI}(k)) &= \left(1 - \sup_{\pi \notin \mathcal{KI}(k)} A_k(\pi) \right) \\ &= 1 - \max \{ 2P_{\pi=\mathcal{U}(k)}(X \geq k), 2P_{\pi=O(k)}(X \leq k) \}. \end{aligned} \quad (4.14)$$

Beweis:

Nach Lemma 4.1.14 ist die 2-seitige p -Wert Funktion stetig. Außerdem verlaufen die Funktionswerte der Untergrenzen monoton steigend, die der Obergrenzen monoton fallend. Demnach tritt $\sup_{\pi \notin \mathcal{KI}(k)} A_k(\pi)$ an $P_{\mathcal{U}(k)}(X \geq k)$ oder an $P_{O(k)}(X \leq k)$ auf, woraus die Behauptung folgt. \square

Beispiel:

Die Berechnung der p -Konfidenz wird an einem Beispiel erläutert. Als Grundlage dient das Bernoulli-Modell (4.1). Es seien X_1, \dots, X_n *u.i.v.* $\mathcal{B}(\pi)$ und $k = \sum_{j=1}^n X_j$ die Anzahl der Erfolge. Zur Berechnung der p -Konfidenz muss explizit ein Konfidenzintervall für eine Erfolgswahrscheinlichkeit betrachtet werden. Die Berechnung wird am sogenannten 'Wald'-Konfidenzintervall gezeigt, welches im Kapitel 4.2 unter (4.20) vorgestellt wird. Werden nun 5 Erfolge in 20 Versuchen beobachtet, also $k = 5$, $n = 20$, dann ergibt sich eine relative Häufigkeit von $\hat{\pi} = \frac{5}{20} = \frac{1}{4}$. Das Signifikanzniveau sei $\alpha = 5\%$. Die Grenzen des Wald-Konfidenzintervalls berechnen sich somit zu

$$\left[\hat{\pi} - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\hat{\pi}(1-\hat{\pi})}, \hat{\pi} + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\hat{\pi}(1-\hat{\pi})} \right] = [0.06, 0.44].$$

Diese Grenzen müssen nun in Satz 4.1.16 eingesetzt werden. Zu berechnen ist also:

$$\begin{aligned} c_n &= 1 - \max \left\{ 2 \left(1 - \sum_{i=0}^4 \binom{20}{i} 0.06^i (1-0.06)^{20-i} \right), 2 \sum_{i=0}^5 \binom{20}{i} 0.44^i (1-0.44)^{20-i} \right\} \\ &= 0.8674462. \end{aligned}$$

Das Wald-Konfidenzintervall hat also in diesem Versuch eine p-Konfidenz von $c_n = 0.87$. Dieses bedeutet, dass der korrespondierende Wald-Test die Nullhypothese $H_0 : \pi = \pi_0$ mit außerhalb des Intervalls liegenden Werten π_0 nicht ablehnt.

Um die p-Konfidenz nun anschließend genauer und allgemeiner interpretieren zu können, wird zunächst ein einfaches Bernoullimodell (4.1) unterstellt mit

$$X_1, X_2, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(\pi) \tag{4.15}$$

und binomialverteilter Zufallsvariable $X = \sum_{j=1}^n X_j \sim \text{BIN}(n, \pi)$.

Wird nun die binomialverteilte Zufallsvariable X durch k beobachtet, dann wird mit k ein Konfidenzintervall $\mathcal{KI}(k)$ für π konstruiert.

Die p-Konfidenz des Konfidenzintervalls $\mathcal{KI}(k)$ sei $c_n(\mathcal{KI}(k))$.

- 1.Fall: Es gelte $c_n(\mathcal{KI}(k)) = (1 - \alpha - \delta)$, mit einem geeigneten $\delta > 0$. Dann gilt: durch die Beobachtung k wird die Hypothese $H_0 : \pi = \pi_0$ für $\pi_0 \notin \mathcal{KI}(k)$ mit einem zum Konfidenzintervall korrespondierenden p-Wert in Höhe von $\alpha + \delta$ nicht abgelehnt.
- 2.Fall: Es gelte $c_n(\mathcal{KI}(k)) = (1 - \alpha + \delta)$, mit einem geeigneten $\delta > 0$. Dann gilt: durch die Beobachtung k wird die Hypothese $H_0 : \pi = \pi_0$ für $\pi_0 \in \mathcal{KI}(k)$, mit einem zum Konfidenzintervall korrespondierenden p-Wert in Höhe von $\alpha - \delta$ abgelehnt.

Definition 4.1.17 (*p*-BIAS)

Es gelte das Modell (4.1) mit $X \sim \text{BIN}(n, \pi)$. Die Beobachtung von X sei k . $A_k(\pi)$ sei die 2-seitige p-Wert Funktion und $[\mathcal{KI}(k)] = [\mathcal{U}(k), \mathcal{O}(k)]$ ein Konfidenzintervall für π . Die Zahl

$$b(\mathcal{KI}(k)) = \max \left\{ 0, \sup_{\pi \notin \mathcal{KI}(k)} A_k(\pi) - \inf_{\pi \in \mathcal{KI}(k)} A_k(\pi) \right\}$$

heißt *b*-Bias des Intervalls \mathcal{KI} .

Auch hier dient das vereinfachte Bernoullimodell (4.1) zur Interpretation des p-Bias. Außerdem wird ein zum Konfidenzintervall \mathcal{KI} assoziierter Test für $H_0 : \pi = \pi_0$ betrachtet.

Der p-Bias eines Intervalls $\mathcal{KI}(k)$ misst die Differenz zwischen den assoziierten p-Werten der Tests für H_0 . Er gibt an, wieviel die p-Werte für Tests der Form $H_0 : \pi = \pi_0$ für $\pi_0 \notin \mathcal{KI}$ größer sein dürfen, als für $\pi_0 \in \mathcal{KI}$. Der p-Bias sollte demnach für alle $k \in \mathbb{N}$ klein sein.

4.1.4 Konfidenzintervalle mit $X=0$ (Null- bzw. Vollergebnis)

Im Fall des Nullergebnisses (unter Verwendung des Modells (4.1)) erhält man ein gutes Konfidenzintervall für π durch die folgende Überlegung:

Für eine binomialverteilte Zufallsvariable X gilt:

$$\begin{aligned} P(X = 0) &= \binom{n}{0} \cdot \pi^0 \cdot (1 - \pi)^{n-0} \\ &= (1 - \pi)^n. \end{aligned}$$

Diese soll dem gegebenen Niveau α entsprechen. Folglich gilt:

$$(1 - \pi)^n \stackrel{!}{=} \alpha \Leftrightarrow \pi = 1 - \sqrt[n]{\alpha}.$$

Damit folgt für $X = 0$:

$$\mathcal{KI}(0) = [0, 1 - \sqrt[n]{\alpha}]. \tag{4.16}$$

Eine analoge Darstellung ergibt sich für das Vollergebnis $X=n$.

4.2 Konfidenzintervalle für Erfolgswahrscheinlichkeit π

Die Konstruktion von Konfidenzintervallen für eine Erfolgswahrscheinlichkeit π ist mit Sicherheit Bestandteil jeder Grundvorlesung über statistische Datenanalyse. Bekannt sind unterschiedliche Vorgehensweisen bei der Konstruktion von Konfidenzbereichen.

Zur Vereinfachenden Darstellung derer wird das Bernoullimodell (vgl. (4.1)) mit

$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{B}(\pi) \Rightarrow X = \sum_{j=1}^n X_j \sim \text{BIN}(n, \pi)$ betrachtet. Die relative Häufigkeit

$$\hat{\pi} = \frac{X}{n} = \frac{1}{n} \sum_{j=1}^n X_j, \text{ wobei } E(\hat{\pi}) = \pi \text{ und } \text{Var}(\hat{\pi}) = \frac{1}{n} \pi(1 - \pi) \quad (4.17)$$

gilt, ist ein erwartungstreuer Schätzer für π . Ausserdem gilt, dass

$$\widehat{\text{Var}(\hat{\pi})} = \frac{\hat{\pi}(1 - \hat{\pi})}{n} \quad (4.18)$$

einen konsistenten Schätzer für $\text{Var}(\hat{\pi})$ bezeichnet.

Beweis:

Der Beweis kann im Lehrbuch 'Statistical Inference' von Rohatgi [3] nachgelesen werden. \square

Die Anwendung des zentralen Grenzwertsatzes von Levy (vgl. A.0.5) und des Satzes von Slutsky (A.0.9) liefert schließlich für die Statistik

$$\frac{\hat{\pi} - \pi}{\sqrt{\widehat{\text{Var}(\hat{\pi})}}} = \sqrt{n} \frac{\hat{\pi} - \pi}{\hat{\pi}(1 - \hat{\pi})} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (4.19)$$

die asymptotische Normalität.

Aus der in (4.19) gezeigten Verteilung der relativen Häufigkeit und durch Schätzung des Standardfehlers läßt sich nun leicht ein Konfidenzintervall für π berechnen. Die Grenzen dieses Intervalls sind gegeben durch

$$\hat{\pi} \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\hat{\pi}(1 - \hat{\pi})} \quad (4.20)$$

und wird in der Literatur als 'Wald'-Konfidenzintervall bezeichnet. Der Ausdruck $z_{1-\alpha/2}$ bezeichnet das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

Der Name dieses Intervalls folgt aus der Tatsache, dass dieses aus der Invertierung eines sogenannten Wald-Tests entsteht. Im allgemeinen unterliegen Wald-Tests der Form

$$\frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \quad (4.21)$$

und beinhalten stets eine Maximum-Likelihood Schätzung $\hat{\theta}$ von θ . $SE(\hat{\theta})$ bezeichnet den Standardfehler des Maximum-Likelihood Schätzers.

Der Vorteil eines Wald-Tests (4.21) liegt darin, dass die asymptotische Verteilung der Statistik bekannt ist, sie konvergiert allerdings nur sehr langsam gegen die Grenzverteilung, was bedeutet, dass dieser Test nur bei sehr hohen Stichprobenumfängen das Niveau einhält. Diese Ergebnisse lassen sich damit direkt auf die Eigenschaften des Wald-Konfidenzintervalls (4.20) übertragen. Dieses hat für kleine Stichprobenumfänge eine sehr geringe Coverage-Probability.

Bezeichnet man weiter die Verteilungsfunktion der binomialverteilten Zufallsvariable X aus (4.1) mit $F_B(k|n, \pi)$, dann gilt

$$F_B(k|n, \pi) = 1 - H\left(\frac{n-k}{k+1} \cdot \frac{\pi}{1-\pi} \middle| 2(k+1), 2(n-k)\right), \quad (4.22)$$

wobei $H(t|f_1, f_2)$ die Verteilungsfunktion der F-Verteilung mit f_1 und f_2 Freiheitsgraden bezeichnet. Betrachtet man weitergehend einen zweiseitigen Hypothesentest für $H_0 : \pi = \pi_0$, dann sind die sogenannten 'exakten' Konfidenzintervalle für π von 'Pearson-Clopper' die Lösungen in π_0 der Gleichungen

$$\begin{aligned} \sum_{j=0}^k \binom{n}{j} \pi_0^j (1 - \pi_0)^{n-j} &= \frac{\alpha}{2} \quad \text{und} \\ \sum_{j=k}^n \binom{n}{j} \pi_0^j (1 - \pi_0)^{n-j} &= \frac{\alpha}{2}, \end{aligned} \quad (4.23)$$

unter der Verwendung von (4.22). Die Grenzen der 'Pearson-Clopper'-Konfidenzintervalle für π sind schließlich gegeben durch

$$\begin{aligned} \mathcal{U}^{PC} &= \frac{k}{k + (n - k + 1) \cdot F_{1-\alpha/2}(f'_1, f'_2)} \quad \text{und} \\ \mathcal{O}^{PC} &= \frac{k + 1}{k + 1(n - k) \cdot F_{1-\alpha/2}(f'_1 - 2, f'_2 + 2)}. \end{aligned} \quad (4.24)$$

Die Freiheitsgrade der F-Verteilungen betragen

$$f'_1 = 2(n - k + 1) \quad \text{bzw.} \quad f'_2 = 2k.$$

$F_{1-\alpha/2}$ sei das $(1 - \alpha/2)$ -Quantil der $F(f'_1, f'_2)$ -Verteilung.

Aus (4.23) folgt direkt, dass die Pearson-Clopper-Intervalle (4.24) eine Coverage-Probability nach Definition 4.1.5 von mindestens $1 - \alpha \forall \alpha \in (0, 1)$ aufweisen. Die mittlere Konservativität \bar{K}_n aus Definition 4.1.9 dieser Intervalle ist demnach sehr hoch, die Anti-Konservativität \overline{AK}_n gleich Null. Die meist als 'Goldstandard' (vgl. [17]) ausgewiesenen Intervalle von Pearson und Clopper haben demnach keine sehr guten Eigenschaften.

Ein weiteres Intervall folgt aus der Invertierung der Wald-Statistik (4.19) zur Hypothese $H_0 : \pi = \pi_0$, dessen Grenzen die Lösungen in π_0 der Gleichungen

$$\sqrt{n} \left(\frac{\hat{\pi} - \pi_0}{\pi_0(1 - \pi_0)} \right) = \pm z_{1-\alpha/2}$$

darstellen. Die Lösungen dieser Gleichungen, 1927 von WILSON publiziert [17], haben die Form

$$\left(\hat{\pi} + \frac{z_{1-\alpha/2}^2}{2n} \pm \sqrt{[\hat{\pi}(1 - \hat{\pi}) + z_{1-\alpha/2}^2/4n]/n} \right) / (1 + z_{1-\alpha/2}^2/n). \quad (4.25)$$

Man nennt dieses Konfidenzintervall für π auch das 'Score'-Konfidenzintervall. Score-Tests basieren auf einer log-likelihood Schätzung unter $H_0 : \pi = \pi_0$ für π und binden damit die Nullhypothese direkt in die Schätzung ein, während 'Wald'-Tests auf einer log-likelihood Schätzung nach der Maximum-Likelihood Methode basieren. Im allgemeinen Fall haben Score-Tests die Form

$$\left(\frac{\partial \log L(\theta)}{\partial \theta} \right)_{H_0: \theta = \theta_0} \geq C$$

mit Likelihood-Funktion $L(\cdot)$. Für weitere Details dieser Schätzverfahren sei auf die Literatur (z.B. AGRESTI (1996) [18], S. 88-95) verwiesen.

Es scheint schwierig das Wilson-Intervall (4.25) und das Wald-Intervall (4.20) zu vergleichen. Das Wald-Intervall ist symmetrisch um den Maximum-Likelihood Schätzer angeordnet, das Wilson-Intervall hingegen besitzt

$$\begin{aligned} \frac{1}{2} \cdot [Q^{Wilson} + O^{Wilson}] &= \hat{\pi} \left(\frac{n}{n + z_{1-\alpha/2}^2} \right) + \frac{1}{2} \left(\frac{z_{1-\alpha/2}^2}{n + z_{1-\alpha/2}^2} \right) \\ &= \frac{1}{n + z_{1-\alpha/2}^2} \left(X + \frac{1}{2} z_{1-\alpha/2}^2 \right) \\ &\approx \frac{1}{n + 4} (X + 2) =: \tilde{\pi} \end{aligned} \quad (4.26)$$

als Mittelpunkt.

Anschaulich bedeutet (4.26), dass sich die geschätzte Erfolgswahrscheinlichkeit nach der Wilson-Methode dadurch auszeichnet, dass diese nicht die gewöhnliche relative Häufigkeit $\hat{\pi} = \frac{X}{n}$ verwendet, sondern sich approximativ wie ein Schätzer verhält, der auf den Gesamtstichprobenumfang vier Ergebnisse hinzuaddiert - zwei Erfolge und zwei Mißerfolge.

AGRESTI und COULL (1998) [17] konstruierten aus diesem Schätzer ein neues Konfidenzintervall für π in der einfachen Form, indem sie diesen Schätzer einfach in die Gleichung des Wald-Konfidenzintervalls (4.20) einsetzten und damit das 'Agresti'-Konfidenzintervall erhielten:

$$\tilde{\pi} \pm \frac{z_{1-\alpha/2}}{\tilde{n}} \sqrt{\tilde{\pi}(1-\tilde{\pi})}, \quad \tilde{n} = n + 4. \quad (4.27)$$

Ein großer Vorteil dieser Methode ist die Einfachheit des Verfahrens. Ein großer Nachteil ist, dass sowohl dieses Intervall, als auch das Wald-Konfidenzintervall, nicht bereichserhaltend sind und über die $[0,1]$ -Begrenzung herausragen können.

Als Motivation zur Konstruktion neuer Intervalle sollen primär diese Grundgedanken dienen:

- Das Konfidenzintervall soll einfach zu berechnen sein.
- Das Konfidenzintervall soll bereichserhaltend sein.

4.3 Transformierte Statistiken

4.3.1 stetige Transformationen

In der Literatur findet man zahlreiche mögliche Transformationsfunktionen um eine Statistik stetig zu modifizieren (vgl. [16]). Nach Kapitel 3.4 kommen alle stetigen und monotonen Bijektionen in Frage, deren erste Ableitungen nullstellenfrei sind.

Satz 4.3.1 (TRANSFORMIERTE STATISTIK)

Es gelte das Modell (4.1) und es seien $X_1, X_2, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(\pi)$. Die Funktion $g: (0,1) \rightarrow \mathbb{R}$ erfülle die Anforderungen an eine Transformationsfunktion aus Kapitel 3.4. Es sei $\hat{\pi}$ die relative Häufigkeit. Dann gilt:

$$\sqrt{n} \left(\frac{g(\hat{\pi}) - g(\pi)}{\sqrt{(\hat{\pi}(1-\hat{\pi})) [g'(\hat{\pi})]^2}} \right) \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0,1).$$

Beweis:

Da $X_1, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(\pi)$ gilt: $E(X_1) = \pi$ und $Var(X_1) = \pi(1-\pi)$. Durch die Anwendung des

zentralen Grenzwertsatzes (A.0.6) gilt nun:

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{L} \mathcal{N}(0, \pi(1 - \pi)).$$

Die Anwendung der δ -Methode (siehe Satz A.0.8) und des Satzes von Slutsky (siehe Satz A.0.9), sowie die Verwendung des konsistenten Schätzers $\widehat{\text{Var}}(\hat{\pi})$ aus (4.18) liefert die Behauptung. \square

Korollar 4.3.2 (TRANSFORMIERTE KONFIDENZINTERVALLE)

Es gelten die Voraussetzungen aus Satz 4.3.1. Ein asymptotisches $(1 - \alpha)$ -Konfidenzintervall für π ist gegeben durch

$$[\mathcal{U}, \mathcal{O}] = [g^{-1}(\mathcal{U}^{Transf}), g^{-1}(\mathcal{O}^{Transf})],$$

wobei

$$\begin{aligned} \mathcal{U}^{Transf} &= g(\hat{\pi}) - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\hat{\pi}(1 - \hat{\pi}) \cdot [g'(\hat{\pi})]^2}, \\ \mathcal{O}^{Transf} &= g(\hat{\pi}) + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\hat{\pi}(1 - \hat{\pi}) \cdot [g'(\hat{\pi})]^2} \end{aligned}$$

gilt. Die Notation $z_{1-\alpha/2}$ bezeichnet dabei das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

Beweis:

Wegen Satz 4.3.1 und aufgrund der Monotonie und der Umkehrbarkeit von g folgt:

$$\begin{aligned} P(\pi \in [\mathcal{U}, \mathcal{O}]) &= P(g(\pi) \in [\mathcal{U}^{Transf}, \mathcal{O}^{Transf}]) \\ &= P\left(z_{\alpha/2} \leq \sqrt{n} \left(\frac{g(\hat{\pi}) - g(\pi)}{\sqrt{(\hat{\pi}(1 - \hat{\pi})) [g'(\hat{\pi})]^2}} \right) \leq z_{1-\alpha/2} \right) \xrightarrow{n \rightarrow \infty} 1 - \alpha. \end{aligned}$$

\square

Um nun die zur Auswahl stehenden Funktionen weiter einschränken zu können, hilft eine besondere Eigenschaft der relativen Häufigkeit: Äquivarianz.

Satz 4.3.3 (ÄQUIVARIANZ DER RELATIVEN HÄUFIGKEIT)

Es gelte das Modell (4.1) und es seien $X_1, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(\pi)$. Die relative Häufigkeit sei $\hat{\pi}$. Dann ist $\hat{\pi}$ äquivariant bezüglich jeder streng monoton steigenden Abbildung ϕ bzw. streng monoton fallender Funktion ϕ^* .

Beweis:

Der Beweis wurde ausführlich in der Arbeit von Domhof [13] für relative Effekte geführt und kann direkt auf Erfolgswahrscheinlichkeiten übertragen werden. \square

Aus Satz (4.3.3) folgt, dass sich die relative Häufigkeit bei einer streng monoton steigenden Transformation der Zufallsvariablen nicht verändert, während sich der Schätzer nach streng monoton fallender Transformation der Zufallsvariablen am Wert $\frac{1}{2}$ spiegelt. Damit sich dieses auf die transformierten Konfidenzintervalle überträgt, kann außerdem von der Transformationsfunktion $g : \Theta \rightarrow \mathbb{R}$ gefordert werden:

(T5) Der Graph von g soll punktsymmetrisch zum Mittelpunkt ξ von Θ verlaufen.

Aus (T1) – (T5) folgt, dass nun ohne Beschränkung der Allgemeinheit Bijektionen mit

$$(T6) \quad g(\xi) = 0$$

in Betracht gezogen werden können. Bekannte Funktionen, die die Eigenschaften (T1) – (T6) erfüllen, sind die sogenannten *Logit*- und *Probit*-Funktionen, die in Lemma 4.3.4 dargestellt sind.

Lemma 4.3.4 (*LOGIT - bzw. PROBIT-KONFIDENZINTERVALL*)

Es gelte das Modell (4.1) und es seien $X_1, \dots, X_n \stackrel{u.i.v.}{\sim} \mathcal{B}(p)$, $p \in (0, 1)$. Die relative Häufigkeit sei $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$. Φ sei die Verteilungsfunktion der Standardnormalverteilung. Betrachtet man weiter die Abbildungen

$$\text{Logit} : (0, 1) \rightarrow \mathbb{R}, \quad p \mapsto \log\left(\frac{p}{1-p}\right), \quad (4.28)$$

mit der Umkehrfunktion

$$\text{Expit} : \mathbb{R} \rightarrow (0, 1), \quad x \mapsto \frac{\exp(x)}{1 + \exp(x)}, \quad (4.29)$$

sowie

$$\text{Probit} : (0, 1) \rightarrow \mathbb{R}, \quad p \mapsto \Phi^{-1}(p), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt, \quad (4.30)$$

dann sind durch

$$\begin{aligned} P(\text{Expit}(\mathcal{U}_{\text{Logit}}(k)) \leq p \leq \text{Expit}(\mathcal{O}_{\text{Logit}}(k))) &\approx 1 - \alpha, \\ [\mathcal{U}_{\text{Logit}}(k), \mathcal{O}_{\text{Logit}}(k)] &= \text{Logit}(\hat{p}) \mp \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{\frac{1}{\hat{p}(1-\hat{p})}}, \end{aligned} \quad (4.31)$$

sowie durch

$$\begin{aligned} P(\Phi(\mathcal{U}_{\text{Probit}}(k)) \leq p \leq \Phi(\mathcal{O}_{\text{Probit}}(k))) &\approx 1 - \alpha, \\ [\mathcal{U}_{\text{Probit}}(k), \mathcal{O}_{\text{Probit}}(k)] &= \text{Probit}(\hat{p}) \mp \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{\sqrt{2\pi}}{\exp\{-\frac{1}{2}\text{Probit}^2(\hat{p})\}} \sqrt{\hat{p}(1-\hat{p})}, \end{aligned} \quad (4.32)$$

zwei bereichserhaltende und asymptotische $(1 - \alpha)$ -Konfidenzintervalle gegeben.

$z_{1-\alpha/2}$ bezeichnet dabei das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

Beweis:

Die beiden Funktionen erfüllen die Anforderungen (T1) – (T4) an eine Transformationsfunktion aus Kapitel 3.4, sowie die Anforderungen (T5) – (T6). Weiterhin gilt:

$$\frac{d(\text{Logit})}{dp} = \frac{1}{p(1-p)}, \quad \frac{d(\text{Probit})}{dp} = \frac{\sqrt{2\pi}}{\exp\{-\frac{1}{2}\text{Probit}^2(p)\}}.$$

Die Behauptung folgt schließlich durch die Anwendung von Korollar 4.3.2. \square

Betrachtet man Funktionenscharen der Form

$$\Phi_{0,\sigma}^{-1} : (0, 1) \rightarrow \mathbb{R}, \quad p \mapsto \Phi_{0,\sigma}^{-1}(p), \quad \text{wobei } \Phi_{0,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^x e^{-\frac{t^2}{2\sigma^2}} dt \text{ gilt,} \quad (4.33)$$

dann erfüllen auch diese die Anforderungen (T1) – (T6). Jedoch sind die aus den Funktionenscharen resultierenden Konfidenzintervalle alle gleich, wie folgendes Lemma zeigt.

Lemma 4.3.5 (NICHTPARAMETRISIERBARKEIT DER PROBIT-TRANSFORMATIONSFUNKTION)

Es gelte das Modell (4.1). Das Konfidenzintervall $\mathcal{K}I_\sigma$ sei das Probit(σ)-Intervall aus Lemma 4.3.4 unter der Verwendung der Funktionenscharen $\Phi_{0,\sigma}(x)$ aus (4.33). Für die Konfidenzintervalle gilt dann stets:

$$\mathcal{K}I_\sigma = [\mathcal{U}_{\text{Probit}(\sigma)}, \mathcal{O}_{\text{Probit}(\sigma)}] = [\mathcal{U}_{\text{Probit}(1)}, \mathcal{O}_{\text{Probit}(1)}].$$

Beweis:

$$\Phi_{0,\sigma} \left[\underbrace{\Phi_{0,\sigma}^{-1}(\hat{p}) \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{\sqrt{2\pi}\sigma}{\exp\left\{-\frac{(\Phi_{0,\sigma}^{-1}(\hat{p}))^2}{2\sigma^2}\right\}} \sqrt{\hat{p}(1-\hat{p})}}_{x(\sigma)} \right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x(\sigma)} e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} dt. \quad (4.34)$$

Substitution liefert schließlich für

$$u = \frac{t}{\sigma} \Rightarrow \frac{du}{dt} = \frac{d}{dt} \left(\frac{t}{\sigma} \right) \Rightarrow dt = du \cdot \sigma$$

und damit gilt:

$$\begin{aligned} (4.34) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{x(\sigma)}{\sigma}} e^{-\frac{1}{2}u^2} du \cdot \sigma = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x(\sigma)}{\sigma}} e^{-\frac{1}{2}u^2} du = \Phi_{0,1} \left(\frac{x(\sigma)}{\sigma} \right) \\ &= \Phi_{0,1} \left[\frac{1}{\sigma} \left\{ \Phi_{0,\sigma}^{-1}(\hat{p}) \pm \frac{\sqrt{2\pi}\sigma}{\exp\left\{-\frac{(\Phi_{0,\sigma}^{-1}(\hat{p}))^2}{2\sigma^2}\right\}} \sqrt{\hat{p}(1-\hat{p})} \right\} \right] \\ &= \Phi_{0,1} \left[\Phi_{0,1}^{-1}(\hat{p}) \pm \frac{\sqrt{2\pi}}{\exp\left\{-\frac{-(\Phi_{0,\sigma}^{-1}(\hat{p}))^2}{2\sigma^2}\right\}} \sqrt{\hat{p}(1-\hat{p})} \right] \\ &= \Phi_{0,1} \left[\Phi_{0,1}^{-1}(\hat{p}) \pm \frac{\sqrt{2\pi}}{\exp\left\{-\frac{-(\Phi_{0,1}^{-1}(\hat{p}))^2}{2}\right\}} \sqrt{\hat{p}(1-\hat{p})} \right] \\ &= \Phi_{0,1}(x(1)). \end{aligned}$$

□

Das Logit-Intervall und Probit-Intervall (4.31 und 4.32) zeigen folgenden Nachteil auf:

- Für $k = 0 \Rightarrow \hat{\pi} = 0$ degenerieren diese Intervalle bzw. sind für Null- oder Vollergebnisse nicht berechenbar.

Diese Gegebenheit ist problematisch, da sich diese Intervalle in der gezeigten Form für Null- bzw. Vollergebnisse nicht berechnen lassen. Der Grundgedanke, eine einheitliche Formel für ein Intervall herzuleiten, soll motivierend dafür sein, das Logit-Intervall zu parametrisieren. Ein weiterer Vorteil durch eine Parametrisierung des Logit-Intervalls liegt darin, dass diese Konfidenzchar

schließlich optimiert werden kann, falls eine analytische Begründung für das Vorliegen eines Optimums gefunden wird. Betrachtet man also Funktionenscharen der Form

$$\text{Logit}_c : (0, 1) \rightarrow \mathbb{R}, \quad \pi \mapsto \log \left(\frac{\pi + c}{1 + c - \pi} \right), \quad (4.35)$$

mit einem gegebenem $c \in (0, 1)$, dann gilt für die erste Ableitung:

$$\frac{d(\text{Logit}_c)}{d\pi} = \frac{1 + 2c}{\pi(1 - \pi) + c(1 + c)}.$$

Diese Funktionenschar ist bijektiv und damit umkehrbar. Als Umkehrabbildung braucht aber nicht die exakte Umkehrabbildungsschar betrachtet werden, denn die Transformationsmethode verlangt eine Funktion, die die offene Menge $\Theta \subset \mathbb{R}$ auf das offene Einheitsintervall abbildet. Als Umkehrabbildung kann daher die *Expit*-Funktion betrachtet werden.

Lemma 4.3.6 (*MODIFIZIERTES LOGIT-INTERVALL*)

Es gelte das Bernoullimodell (4.1). Weiter wird die Funktionenschar der Logit_c -Transformation aus (4.35) mit *Expit*-Funktion als Umkehrung (4.29) betrachtet². Dann sind durch

$$P(\text{Expit}(\mathcal{U}_{\text{Logit}_c}^*) \leq \pi \leq \text{Expit}(O_{\text{Logit}_c}^*)) = 1 - \alpha,$$

wobei

$$[\mathcal{U}_{\text{Logit}_c}^*, O_{\text{Logit}_c}^*] = \text{Logit}_c(\hat{\pi}) \mp \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{1 + 2c}{\hat{\pi}(1 - \hat{\pi}) + c(1 + c)} \sqrt{\hat{\pi}(1 - \hat{\pi})} \quad (4.36)$$

gilt, asymptotische und bereichserhaltende $(1 - \alpha)$ -Konfidenzintervalle für π definiert.

Beweis:

Die Transformationsfunktionen erfüllen für $0 < c < 1$ die Anforderungen (T1) – (T4) sowie (T5) – (T6). Die Behauptung folgt direkt durch die Anwendung von Korollar 4.3.2. \square

Hergeleitet wurden nun Konfidenzscharen für eine Erfolgswahrscheinlichkeit. Das Problem ist nun, dass ein bestimmtes Intervall aus dieser Schar ausgewählt werden muss. Diese Auswahl muss analytisch begründet werden und kann nicht willkürlich erfolgen. Aufgrund der Unstetigkeit der Coverage-Prozesse ist dieses schwierig. Allerdings ist die Längenkurve eine stetige Abbildung und damit mit analytischen Methoden berechenbar. Es wird daher so vorgegangen und untersucht, wie sich die mittlere Länge der Konfidenzchar bei wachsender Transformationskonstante verhält. Dieses Verhalten wird im folgenden Satz 4.3.7 beschrieben. Zunächst wird die Aussage dieses Satzes explizit dargestellt und erst nach dem Beweis dieses Satzes wird die wichtige Folgerung aus diesem erläutert.

Es seien $L_{n,c} : (0, 1) \rightarrow [0, 1]$ die Längenkurve und $\bar{L}_{n,c}$ die mittlere Länge des Logit_c -Intervalls. Im Satz 4.3.7 wird behauptet, dass die Abbildung

$$\Psi : (0, 1) \rightarrow [0, 1], \quad c \mapsto \bar{L}_{n,c}, \quad (4.37)$$

welche jedem $c \in (0, 1)$ die mittlere Länge des Logit_c -Intervalls zuordnet, streng monoton fällt.

²Aus Gründen der Einheitlichkeit wird im folgenden die Erfolgswahrscheinlichkeit wieder mit π bezeichnet

Satz 4.3.7 (MONOTONIE DER MITTLEREN LÄNGEN DES LOGIT_c-INTERVALLS)

Es gelte das Modell (4.1) und $\mathcal{K}I_c(X)$ sei das parametrisierte Logit_c-Intervall aus (4.36) für π . Es sei außerdem $L_{n,c}(\pi)$ das Bild der Längenkurve des Konfidenzintervalls $\mathcal{K}I_c(X)$. Dann ist durch

$$\Psi(c) = \int_0^1 L_{n,c}(\pi) d\pi$$

eine streng monoton fallende Funktion definiert.

Beweis:

Die Funktion Ψ ist nach Lemma 4.1.3 stetig und differenzierbar. Weiter ist die Transformationsfunktion bijektiv und streng monoton steigend. Demnach kann das Intervall im transformierten System betrachtet werden. Bekanntlich gilt für die erwartete Länge des Logit_c-Intervalls (vgl. Lemma 4.1.2):

$$L_{n,c}(\pi) = \sum_{k=0}^n [O_c(k) - \mathcal{U}_c(k)] \cdot \binom{n}{k} \pi^k (1-\pi)^{n-k}. \quad (4.38)$$

Da die erwartete Länge stets größer oder gleich Null ist, kann der Satz folgendermaßen bewiesen werden:

- Fällt die erwartete Länge mit steigender Konstante, dann fällt auch die mittlere Länge.

Zum Beweis des Satzes genügt es daher den Summanden $\mathcal{K}I_c(i)$ aus (4.38) zu betrachten. Es kann auch angenommen werden, dass $(i) \neq (0)$ gilt.

Damit folgt:

$$\begin{aligned} & O_c(i) - \mathcal{U}_c(i) \\ = & \text{Logit}_c(\hat{\pi}) + \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{1+2c}{\hat{\pi}(1-\hat{\pi}) + c(1+c)} \sqrt{\hat{\pi}(1-\hat{\pi})} \\ - & \left[\text{Logit}_c(\hat{\pi}) - \frac{z_{1-\alpha/2}}{\sqrt{n}} \frac{1+2c}{\hat{\pi}(1-\hat{\pi}) + c(1+c)} \sqrt{\hat{\pi}(1-\hat{\pi})} \right] \\ = & 2 \cdot \underbrace{\frac{1+2c}{\hat{\pi}(1-\hat{\pi}) + c(1+c)}}_{=:A} \underbrace{\frac{z_{1-\alpha/2}}{\sqrt{n}}}_{=:K} \underbrace{\sqrt{\hat{\pi}(1-\hat{\pi})}}_{=:V}. \end{aligned}$$

Man kann diesen Ausdruck als differenzierbare Funktion von c auffassen. Für die partielle Ableitung nach c gilt:

$$\frac{\partial(\cdot)}{\partial c} = -4KV \left[\frac{c^2 + c + \frac{1}{2} - (\hat{\pi}(1-\hat{\pi}))}{(\hat{\pi}(1-\hat{\pi}) + c + c^2)^2} \right] < 0,$$

womit die Behauptung bewiesen ist. □

Die Konsequenz dieses Satzes ist, dass sich die Parametrisierungskonstante c der Logit-Transformationsfunktion unmittelbar auf die Coverage-Probability und damit auch direkt auf die beschriebenen Streumaße der Coverage-Prozesse des Logit_c-Intervalls auswirkt. Damit sind diese Konfidenzscharen optimierbar, indem die Streumaße (vgl. die Definitionen 4.1.10 bzw. 4.1.11) in Abhängigkeit von c minimiert werden.

Resultat 4.3.8 (OPTIMALE TRANSFORMATIONSFUNKTION)

Es gelte das Bernoullimodell (4.1) und $\mathcal{K}I_c(X)$ sei das parametrisierte Logit_c-Intervall. Die

Abbildung³ $C_{n,c} : (0, 1) \rightarrow [0, 1]$ sei ein Coverage-Prozess des Konfidenzintervalls $\mathcal{KI}_c(X)$. Die Größen $\bar{F}_{n,c}$ bzw. $\bar{S}_{n,c}$ seien der mittlere Fehler bzw. die mittlere Streuung (vgl. die Definitionen (4.1.10) bzw. (4.1.11)) von $\mathcal{KI}_c(X)$ in Abhängigkeit des Stichprobenumfanges n und der Parametrisierungskonstante c des Logit-Intervalls. Dann führt

$$c^* = \frac{1}{150}$$

zu einer mittleren minimalen Schwankung des Coverage-Prozesses um das $(1 - \alpha) = 0.95$ Konfidenzniveau.

Herleitung:

Zu betrachten ist ein Optimierungsmodell aus dem Bereich der eindimensionalen Optimierung. Satz 4.3.7 sichert die Existenz lokaler Minima der folgenden Zielfunktionen.

Die zu optimierende Zielfunktion lautet in allgemeiner Form:

$$Z(z) \rightarrow \min_z, \quad (4.39)$$

unter der Nebenbedingung:

$$z > 0 \quad (\text{Nichtnegativitätsbedingung}). \quad (4.40)$$

Natürlich muss der Optimierungsalgorithmus mehrere Stichprobenumfänge enthalten. Angewendet auf (4.39) können folgende Optimierungstableaus berechnet werden:

$$Z_1(c) = \frac{1}{N} \sum_{j=1}^N \bar{F}_{n_j, c} \rightarrow \min_c, \quad (4.41)$$

$$Z_2(c) = \frac{1}{N} \sum_{j=1}^N \bar{S}_{n_j, c} \rightarrow \min_c, \quad (4.42)$$

unter der Nebenbedingung:

$$0 < c < 1.$$

$N = \#\{n_j\}$ sei die Anzahl der Stichprobenumfänge.

Vorgehen:

Der Optimierungsalgorithmus (4.39) verlangt ein Suchverfahren über einer breiten Auswahl an Stützstellen.

Als Stützstellen dienen folgende Stichprobenumfänge:

- **Stützstellen:**

$$n_j = 5, 15, 20, 30, 50, 100.$$

Die numerischen Integrationen werden mit dem Verfahren aus (3.3) durchgeführt.

Quellcode:

Abschließend wird der Quellcode zur Ermittlung der Kombination gezeigt. Diese Optimierung ist auch im Programmpaket 'R' durchführbar. Die selbst geschriebene Funktion 'Cov.logit' ermittelt die exakte Coverage-Probability und ist hier jetzt nicht gezeigt.

³Die Aufnahme der Konstanten c in den Index bedeutet, dass auf diesem Parameter bedingt wird.

```

nmin<-function(nvec,p,kopt,vopt,aopt)
{
  # Funktion für mehrere Stichprobenumfänge
  ln<-length(nvec)
  erg<-matrix(0,ncol=5,nrow=ln)
  ergebnissumme<-c(rep(0,5))
  for (i in 1:ln){
    hilf<-cov.logit(p,nvec[i],kopt,vopt,aopt) # Berechne exakte Coverage-Probability
    erg[i,1]<-hilf[1]
    erg[i,2]<-hilf[2]
    erg[i,3]<-hilf[3]
    erg[i,4]<-hilf[4]
    erg[i,5]<-hilf[5]
  }
  ergebnissumme[1]<-sum(erg[,1])/ln
  ergebnissumme[2]<-sum(erg[,2])/ln
  ergebnissumme[3]<-sum(erg[,3])/ln
  ergebnissumme[4]<-sum(erg[,4])/ln
  ergebnissumme[5]<-sum(erg[,5])/ln
  ergebnissumme
}
optk<-function(nvec,p,kvec,vopt,aopt){
  lk<-length(kvec)
  ln<-length(nvec)
  ergcov<-c(rep(0,lk))
  ergl<-c(rep(0,lk))
  ergfe<-c(rep(0,lk))
  ergst<-c(rep(0,lk))
  ergvar<-c(rep(0,lk))
  for (i in 1:lk){
    kopt<-kvec[i]
    hilf1<-nmin(nvec,p,kopt,vopt,aopt)
    ergcov[i]<-hilf1[1]
    ergl[i]<-hilf1[2]
    ergfe[i]<-hilf1[3]
    ergst[i]<-hilf1[4]
    ergvar[i]<-hilf1[5]
  }
  laenge<-c(ergl)
  coverage<-c(ergcov)
  fehler<-c(ergfe)
  streu<-c(ergst)
  #print(ergcov)
  #print(ergl)
  #print(ergfe)
  #print(ergst)
  #print(ergvar)
  erg<-matrix(c(fehler,streu,laenge),ncol=3,nrow=lk)
  erg
}

```

Ergebnis:

Das Ergebnis des Suchverfahrens wird grafisch in Abbildung 4.3 gezeigt. In dieser ist sowohl das Bild der Zielfunktionen Z_1 bzw. Z_2 , als auch die streng monoton fallende Längenkurve dargestellt.

In Abbildung 4.3 ist deutlich zu sehen, dass sich die Längenkurve an der Stelle $c = 1/150$ krümmt. Abschließend wird der Verlauf der vorgestellten Transformationsfunktionen grafisch in Abbildung 4.4 gezeigt. In dieser ist außerdem die parametrisierte Logit-Funktion zum Wert $c = 1/150$ dargestellt. Man erkennt im Verlauf dieser Funktion kaum einen Unterschied zur nichtparametrisierten Variante. Betrachtet man den Rand der Funktionen, dann zeigt dieser, dass sich dort die Funktionsgraphen splitten. Die eingebaute Konstante wirkt sich daher primär auf kleine bzw. hohe Effekte aus.

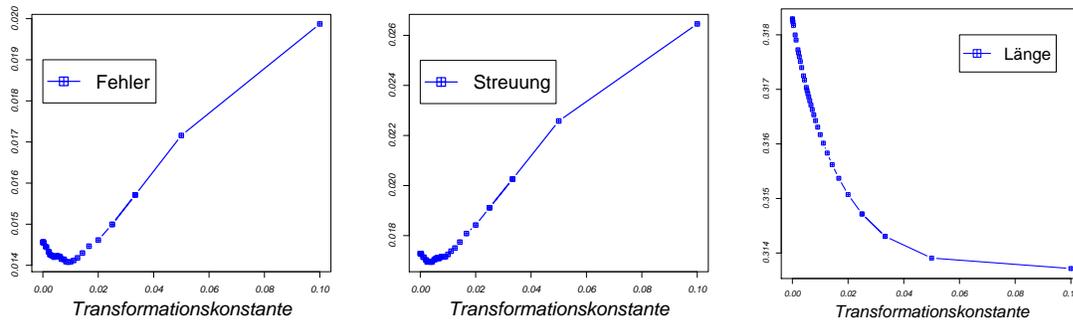


Abbildung 4.3: Ergebnisse der c-Optimierung

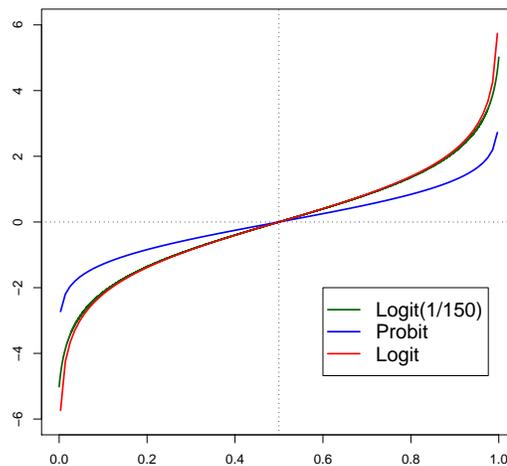


Abbildung 4.4: Logit- und Probit- Transformationsfunktion

4.4 Vergleich der Konfidenzintervalle

4.4.1 Coverage-Prozesse

Im ersten Schritt des Vergleichs der vorgestellten Konfidenzintervalle werden die Coverage-Prozesse grafisch analysiert. Programmiert wurden die Grafiken im Programmpaket 'R' mit Hilfe von Matrizen und Klammeroperatoren, d.h., es wurden Matrizen mit den Einträgen der jeweiligen Ober- bzw. Untergrenzen angelegt und schließlich mit Hilfe von Klammeroperatoren die Indikatoren berechnet. Die Prozesse zeigen die Abbildungen 4.5 und 4.6 als Funktion von $p \in (0, 1)$. Bei der Anfertigung der Arbeit zeigte sich, dass bei geringen Stichprobenumfängen (z.B. $n = 5, 10$) die speziellen Intervalle aus Kapitel(4.1.4) bei auftretenden Null- und Vollergebnissen besser als die hergeleiteten sind. In der Literatur werden die Konfidenzintervalle für diese geringen Stichprobenumfänge untersucht und verglichen (vgl. AGRESTI und COULL (1998) [17]). Aus diesem Grund werden in dieser Arbeit die Coverage-Prozesse für sehr geringe Stichprobenumfänge gezeigt und die Grenzen des Logit- bzw. Probit-Intervalls bei Null- und Vollergebnissen gemäß 4.16

abgefangen. Bei höheren Stichprobenumfängen zeigt das Logit Intervall auch ohne abgefangene Grenzen gute Ergebnisse. Das Logit-Intervall bezeichnet hier das modifizierte $Logit_c$ -Intervall mit $c = 1/150$.

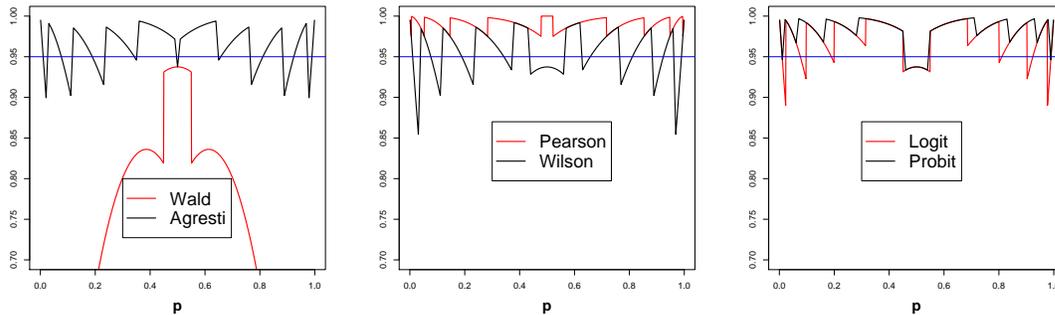


Abbildung 4.5: Coverage-Kurven für $n=5$

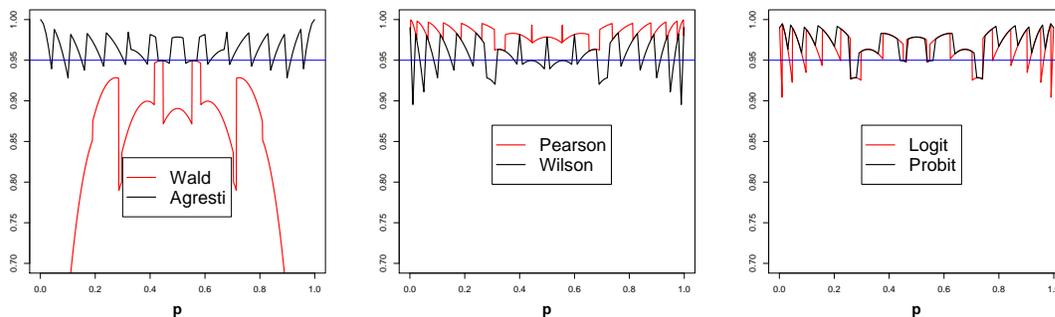


Abbildung 4.6: Coverage-Kurven für $n=10$

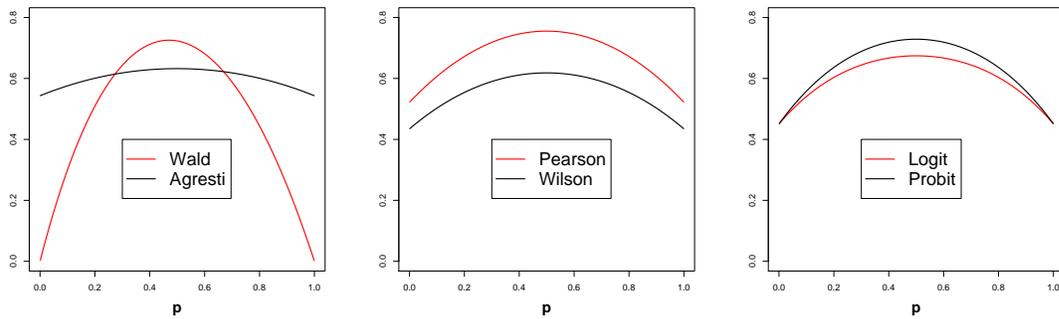
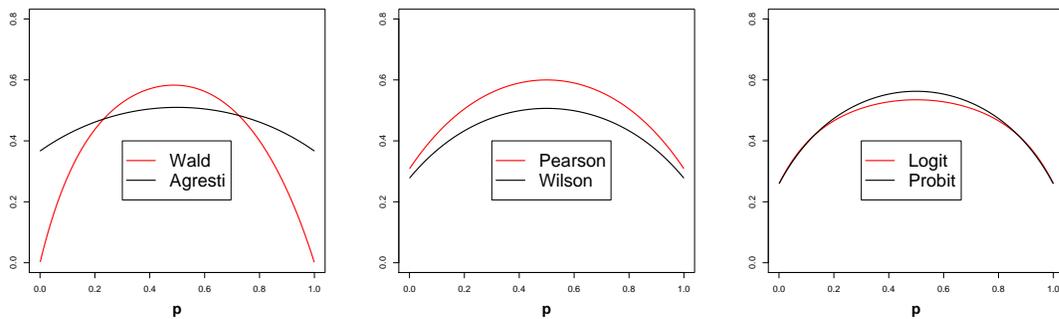
In den Abbildungen 4.5 und 4.6 sind die Coverage-Kurven der vorgestellten Konfidenzintervalle für geringe Stichprobenumfänge dargestellt. Auffällig ist die schlechte Coverage-Probability des Wald-Konfidenzintervalls. Gerade wenn p nahe Null oder Eins ist, scheint dieses Konfidenzintervall unangebracht zu sein. Die schlechte Coverage-Probability am Rand des Wald-Intervalls ist eine Folgerung aus der Punktschätzung für π durch $\hat{\pi}$, denn die Binomialverteilung ist für kleine bzw. große Erfolgswahrscheinlichkeiten schief.

Das Pearson-Clopper-Intervall zeichnet sich durch starke Konservativität bei geringen Stichprobenumfängen aus. Das Agresti- und das Wilson-Intervall schwanken um das Nominalniveau von hier 95 %. Ähnlich sehr gute Ergebnisse zeigen die Coverage-Kurven des Logit- bzw. Probit-Intervalls. Für $n=5$ zeigt das Logit-Intervall eine fast identische Coverage-Kurve zum Wilson-Intervall. Das Probit-Intervall ist leicht konservativ. Für kleine bzw. große p haben diese Intervalle eine sehr gute Coverage-Probability.

Zum Vergleich dazu sollen nun die Längenkurven der Intervalle betrachtet werden.

4.4.2 Ergebnisse numerischer Integrationen

Als numerische Integrationsmethode wurde die vorgestellte Trapezregel aus Kapitel 3.3 angewendet. Um die numerischen Werte der mittleren Coverage, mittleren Konservativität bzw. Anti-

Abbildung 4.7: Längen-Kurven für $n=5$ Abbildung 4.8: Längen-Kurven für $n=10$

Konservativität und Länge zu erhalten, wurde das $(0,1)$ -Intervall der Abszisse in 10.000 Einzelintervalle zerlegt und in diesen die Flächen unterhalb der Kurven approximiert. Zur Berechnung der neuen Maße der Konservativität bzw. Anti-Konservativität wurden mit Hilfe von sogenannten Logik-Operatoren (hier: Klammern) die entsprechenden Werte, die der Bedingungsregel genügen, herausgefiltert, und dann das obige Verfahren angewendet. Die Streumaße wurden in analoger Weise durch einsetzen der numerischen Werte berechnet.

Zu einer aussagekräftigen Beurteilung der Konfidenzintervalle anhand dieser Maße sind mehrere Stichprobenumfänge nötig. Um sowohl asymptotische, als auch Resultate für sehr geringe Stichprobenumfänge zu erhalten, wurden numerische Integrationen für Stichprobenumfänge von

$$n_i = 10 \quad 20 \quad 30 \quad 40 \quad 50$$

durchgeführt.

Um das Lesen der Arbeit angenehmer zu gestalten, werden die Ergebnisse graphisch in Abbildung 4.9 dargestellt, in welcher jeweils der Stichprobenumfang gegen die numerische Größe aufgetragen wurde.

Außerdem wurden aus Gründen der Übersichtlichkeit nicht alle vorgestellten Intervalle in die genaue Analyse einbezogen. Die Auswahl erfolgt auf die in der Literatur als sehr gut ausgewiesenen und auf die neu hergeleiteten Verfahren.

Überraschend ist, dass sich trotz der großen Unterschiede in den mittleren Coverage-Probabilities die mittleren Längen der Intervalle kaum unterscheiden. Obwohl z.B. das Wilson-Intervall eine viel größere Coverage-Probability als das Wald-Intervall für geringe Stichprobenumfänge aufweist, ist dieses nicht wesentlich breiter. Betrachtet man die zugehörigen Längenkurven, dann

zeigt sich, dass das Wilson-Intervall, außer an den Rändern, sogar kürzer ist. Gravierend schlechte Ergebnisse zeigen das Wald- und das Pearson-Intervall im Bezug auf die Coverage-Probability, was sich auf die anderen Maße überträgt.

Das Probit-Intervall ist für kleine Stichprobenumfänge leicht konservativ. Im Bezug auf die Streumaße zeigt sich, dass das Wilson- und das Logit-Intervall fast identische Ergebnisse aufweisen und minimal um das Nominalniveau von 95% streuen. Dementsprechend gering ist auch die mittlere Konservativität beim Logit- und Wilson-Intervall. Eine etwas größere Streuung als bei diesen beiden Intervallen wird durch das Agresti-Intervall erzeugt.

Anhand der untersuchten Maße und durch die breit gefächerten Stichprobenumfänge lässt sich zusammenfassend sagen, dass durch die Anwendung der δ -Methode mit der $Logit_c$ -Transformation ein Konfidenzintervall konstruiert wurde, welches durch die Beurteilung anhand von a-priori-Maßen besser als das Agresti-Intervall ist und nahezu fast identische Ergebnisse zum Wilson-Intervall aufweist. Die Vorteile des Logit-Intervalls zum Wilson-Intervall liegen darin, dass dieses leichter zu berechnen ist. Auch die Herleitung ist leichter zu verstehen als die Konstruktion des Wilson-Intervalls, da das Wilson-Intervall durch die Invertierung eines Score-Tests entsteht. Das Logit-Intervall stellt hingegen einen speziellen invertierten Wald-Test dar.

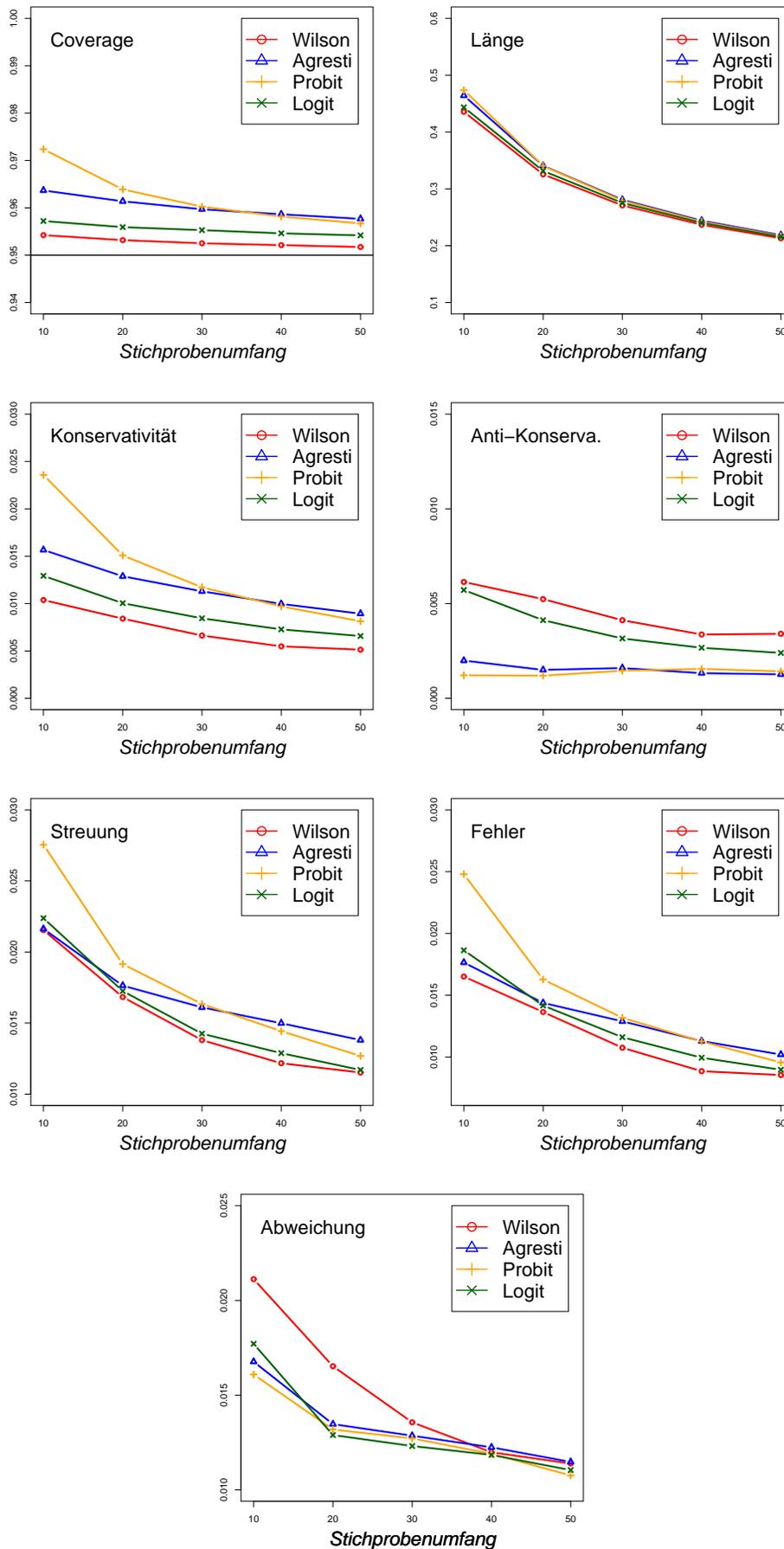


Abbildung 4.9: Ergebnisse numerischer Integrationen für 5 Stichprobenumfänge

4.4.3 p-Konfidenz und p-Bias

Die vorgestellten Konfidenzintervalle wurden anhand der a-posteriori-Maße p-Konfidenz und p-Bias zum 5%-Niveau untersucht. Satz 4.1.16 (Berechnung der p-Konfidenz) erleichtert dabei die Programmierung dieser Maße. Auch hier wurden die Maße in Matrizen-Technik programmiert, d.h., es wurden Matrizen mit Einträgen aus den Ober- bzw. Untergrenzen der Konfidenzintervalle erzeugt und mit Hilfe dieser schließlich die 2-seitige p-Wert Funktion berechnet und anschließend Satz 4.1.16 zur Berechnung dieser Größen hierauf angewendet.

Die Ergebnisse sind grafisch in den Abbildungen 4.10 bis 4.13 für $0 \leq k \leq 10$ bzw. $0 \leq k \leq 50$ dargestellt. Die Maße wurden auch für mehrere Stichprobenumfänge untersucht, bei denen sich immer ähnliche Kurvenverläufe zeigten.

Um die Verfahren gut vergleichen zu können, werden je zwei Intervalle in einer Grafik gezeigt. Aus Gründen der Einheitlichkeit (vgl. Kapitel (4.4.1)) wurden das Logit- bzw. Probit-Intervall im Fall des Null- oder Vollergebnisses durch (4.16) abgefangen. Dieses geschieht aufgrund des geringen Stichprobenumfanges, bei dem die speziellen Intervalle aus (4.16) einfach besser als die neu hergeleiteten sind. Bei einem Stichprobenumfang von $n = 50$ wäre dieses nicht notwendig. Das Logit-Intervall zeigt bei einem Stichprobenumfang von $n = 50$ ohne abgefangene Grenzen ähnlich gute Ergebnisse. Das Logit-Intervall bezeichnet hier das mit $c = 1/150$ parametrisierte Intervall gemäß (4.36).

Aus der Konstruktion der Pearson-Clopper Intervalle folgt sofort, dass diese eine p-Konfidenz von genau $(1 - \alpha) \forall k \in \mathbb{N}$ aufweisen. Der zum Pearson-Clopper-Intervall assoziierte Test für $H_0 : \pi = \pi_0$ führt demnach für $\pi_0 \in [U^{PC}, O^{PC}]$ zu höheren assoziierten p-Werten als Tests für H_0 mit $\pi_0 \notin [U^{PC}, O^{PC}]$. Der p-Bias dieser Intervalle ist demnach gleich Null $\forall k \in \mathbb{N}$. Eine auffällig gute p-Konfidenz zeigt das Probit-Intervall. Selbst bei einem geringen Stichprobenumfang von $n = 10$ hebt sich die p-Konfidenz dieses Intervalls deutlich von der p-Konfidenz der anderen Intervalle ab. Bei einem Stichprobenumfang von $n = 50$ ist ein ähnlich guter Verlauf zu erkennen. Wie bei den Coverage- und Längen-Kurven zeichnen sich das Wilson- und Logit-Intervall durch fast identische Kurvenverläufe aus. Das 95%-Wilson-Intervall zeigt beispielsweise eine p-Konfidenz von weniger als 70% für $n=10$ auf. Dieses bedeutet, dass außerhalb des Intervalls liegende Werte π_0 bei einem korrespondierenden Hypothesentest $H_0 : \pi = \pi_0$ zu einem p-Wert größer als 30% führen. Damit wird die Nullhypothese $H_0 : \pi = \pi_0$ mit nicht im Intervall liegenden Werten π_0 mit dem 'Wilson-Test' nicht abgelehnt, obwohl, nach der Definition eines Konfidenzintervalls, die Nullhypothese verworfen werden müsste.

Weiterhin fällt auf, dass das Agresti-Intervall am Rand eine p-Konfidenz von mehr als 95% aufweist. Analog bedeutet diese Gegebenheit, dass für Werte $\pi_0 \in [U^{Agresti}, O^{Agresti}]$ der korrespondierende Hypothesentest für $H_0 : \pi = \pi_0$ einen assoziierten p-Wert von weniger als 5% aufweist und damit die Nullhypothese ablehnt.

Das Wald-Intervall zeigt bei den beiden dargestellten Stichprobenumfängen den höchsten p-Bias auf. Das Probit-Intervall liefert für beide Stichprobenumfänge einen geringen p-Bias und damit ein sehr gutes Ergebnis. Es zeigt sich, dass das Probit- und Logit-Intervall deutlich besser als das Agresti-Intervall ist.

Zusammenfassend zeigte die Analyse der beiden a-posteriori-Maße, dass die neuen Probit- und Logit-Intervalle nach dem Pearson-Clopper Intervall die besten Intervalle aus den hier untersuchten Intervallen sind.

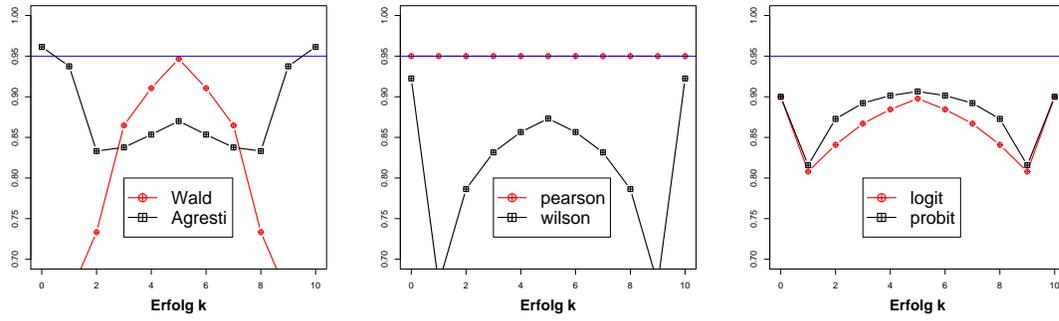


Abbildung 4.10: p-Konfidenz für n=10

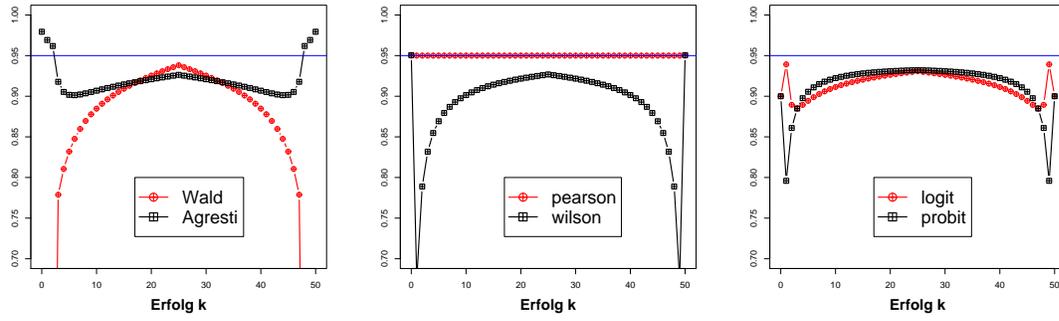


Abbildung 4.11: p-Konfidenz für n=50

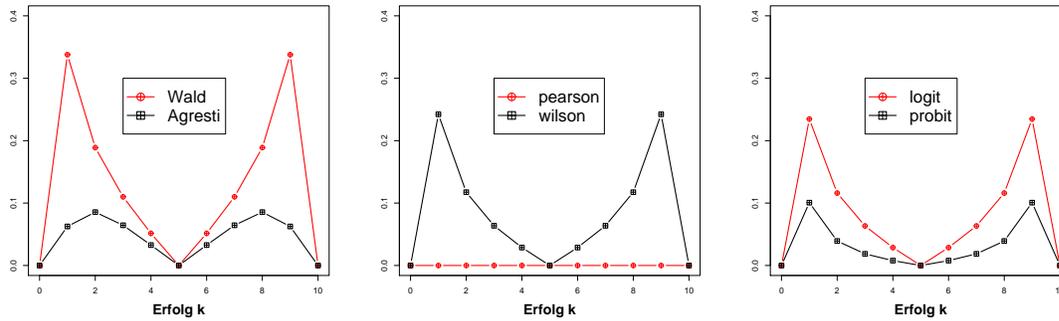


Abbildung 4.12: p-Bias für n=10

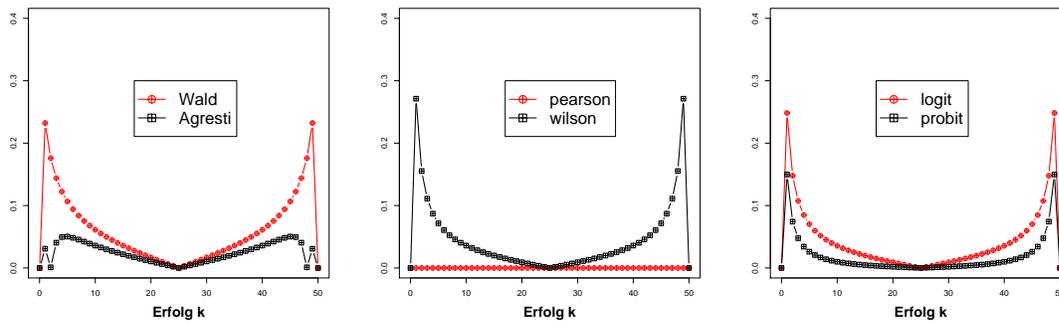


Abbildung 4.13: p-Bias für n=50

4.5 Zusammenfassung und Motivation für Übertragung auf Linearkombinationen

Untersucht wurden unterschiedliche Verfahren zur Konstruktion von Konfidenzintervallen für eine Erfolgsrate π . Neben einer ausführlichen Darstellung von Maßen und Größen zur Bewertung der Intervalle sind neue Konstruktionsverfahren mit Hilfe der Transformationsmethode dargestellt worden. Daneben sind zwei explizite Intervalle durch dieses Verfahren angegeben. Bei der Anfertigung der Arbeit zeigten diese beiden Intervalle mit Hilfe der Logit- und Probit- Transformationsfunktion gute Ergebnisse. Auf diesen Ergebnissen aufbauend sollten die Funktionen parametrisiert werden, um schließlich Konfidenzscharen zu erhalten. Dieses ist schwierig, da sich Paramterisierungskonstanten durch die Rücktransformation häufig gegenseitig aufheben. Außerdem wird die Stetigkeitsforderung verletzt, wenn einzelne Effekte exponiert werden. Die Idee, den Effekt durch Addition einer Konstanten zu verändern und diesen schließlich zu transformieren, tauchte bei Niveausimulationen zu den korrespondierenden Test-Statistiken des Logit- und Probit- Intervalls (in der unparametrisierten Version) auf, welche spezielle Wald-Tests darstellen. Als Schätzer für die Erfolgswahrscheinlichkeiten diente stets die relative Häufigkeit und somit musste bei Null- und Vollergebnissen die Teststatistik abgefangen werden, da diese sonst nicht berechenbar ist.

Die Idee, die Transformation schließlich durch den Einbau einer Konstanten zu parametrisieren (und damit abzufangen) wurde anschließend auf die korrespondierenden Intervalle übertragen. Mit diesem Ergebnis wurde das Verhalten der mittleren Länge der Konfidenzscharen aus der parametrisierten Logit-Funktion betrachtet und bewiesen, dass die mittlere Länge mit wachsender Transformationskonstante streng monoton fällt. Dieses Ergebnis sichert die Existenz einer optimalen Funktion. Das aus einem einfachen Optimierungsalgorithmus resultierende Intervall wurde schließlich mit den anderen Verfahren verglichen und es zeigte sich, dass dieses gut abschneidet. Die Ergebnisse sind motivierend, diese Konstruktionsmethode auch auf Linearkombinationen von Erfolgswahrscheinlichkeiten zu übertragen.

5 Konfidenzintervalle für $p_1 - p_2$ (verbundene Stichproben)

Untersucht und hergeleitet werden nun Konfidenzintervalle für die Differenz zweier Erfolgswahrscheinlichkeiten von verbundenen Stichproben. Als Grundlage dazu dient ein Modell für zwei verbundene Stichproben mit dichotomen Variablen. Aus Gründen der Übersichtlichkeit werden in Zweistichprobendesigns die Erfolgswahrscheinlichkeiten mit $p_j, j = 1, 2$, klassifiziert.

5.1 Modell für zwei verbundene Stichproben

Man betrachte:

$$\begin{aligned} \mathbf{X}_k &= (X_{1k}, X_{2k})', & k &= 1, \dots, n, \\ X_{jk} &\stackrel{u.i.v.}{\sim} \mathcal{B}(p_j), & j &= 1, 2. \end{aligned} \quad (5.1)$$

Damit erhält man folgende Vierfeldertafel mit Erfolgswahrscheinlichkeiten $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$:

		Methode 2		
		1	0	Total
Methode 1	1	$n_{11} (\pi_{11})$	$n_{12} (\pi_{12})$	$n_{11} + n_{12} (p_1)$
	0	$n_{21} (\pi_{21})$	$n_{22} (\pi_{22})$	
Total		$n_{11} + n_{21} (p_2)$		n

Für den Stichprobenumfang gilt dabei $\sum_{i=1}^2 \sum_{j=1}^2 n_{ij} = n$. Als Schätzer für die Raten werden die gewöhnlichen relativen Häufigkeiten verwendet (vgl. das Modell (4.1)).

Definition 5.1.1 (ODDS)

Es seien $\mathbf{X}_k = (X_{1k}, X_{2k})', k = 1, \dots, n, X_{jk} \sim \mathcal{B}(p_j), j = 1, 2$, aus Modell (5.1). Die Maßzahl

$$v_j = \frac{p_j}{1 - p_j}$$

heißt Odds.

Definition 5.1.2 (ODDS-RATIO)

Es gelte das Modell (5.1). Die Maßzahl

$$\theta = \frac{v_1}{v_2}$$

heißt Odds-Ratio.

Der Odds lässt sich dabei als eine Erfolgchance unter der Bedingung bzw. Methode j interpretieren, da dieser die Wahrscheinlichkeit auf Erfolg ins Verhältnis zur Gegenwahrscheinlichkeit setzt. Das Odds-Ratio (Verhältnis von Odds) ist demnach eine relative Erfolgchance unter der Bedingung 1 bezogen auf die Bedingung 2. Liefert das Odds-Ratio einen Wert von 1, dann bedeutet dieses, dass zwischen den Bedingungen 1 und 2 ein gleiches Chancenverhältnis besteht. Die Stichproben sind damit unabhängig.

Weiterhin lässt sich die Kovarianz von $\mathbf{X}_k = (X_{1k}, X_{2k})'$ berechnen mit:

$$\begin{aligned} \text{Cov}(X_{1k}, X_{2k}) &= E(X_{1k}X_{2k}) - E(X_{1k})E(X_{2k}) \\ &= \pi_{11} - p_1p_2 \\ &= \pi_{11} - (\pi_{11} + \pi_{12})(\pi_{11} + \pi_{21}) \\ &= \pi_{11}(1 - \pi_{21} - \pi_{11}) - \pi_{12}(\pi_{11} + \pi_{21}) \\ &= \pi_{11}(\pi_{12} + \pi_{22}) - \pi_{12}(\pi_{11} + \pi_{21}) \\ &= \pi_{11}\pi_{22} - \pi_{12}\pi_{21}. \end{aligned}$$

Die Kovarianz der Zufallsvariablen X_{1k} und X_{2k} ist damit die Differenz der Produkte der Diagonaleinträge der Vierfeldertafel aus Modell (5.1).

Die erwartungstreuen Schätzer für die Erfolgswahrscheinlichkeiten sind gegeben durch:

$$\begin{aligned} \hat{\pi}_{11} &= \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j}X_{2j}, \\ \hat{\pi}_{12} &= \hat{p}_1 - \hat{\pi}_{11} = \frac{1}{n_1} \left(\sum_{k=1}^{n_1} X_{01k}(1 - X_{02k}) \right), \\ \hat{\pi}_{21} &= \hat{p}_2 - \hat{\pi}_{11} = \frac{1}{n_1} \left(\sum_{k=1}^{n_1} X_{02k}(1 - X_{01k}) \right), \\ \hat{\pi}_{22} &= 1 - \hat{\pi}_{11} - \hat{\pi}_{12} - \hat{\pi}_{21}. \end{aligned}$$

Weiterhin beobachtet man, dass der Vektor

$$\mathbf{p}^* = (\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{22})'$$

multinomialverteilt ist, denn es gilt:

- Die Komponente $\hat{\pi}_{ij}$, $i, j = 1, 2$ des Vektors \mathbf{p}^* ist binomialverteilt zu n und π_{ij} .
- Für die Kovarianz der Komponenten $\hat{\pi}_{ij}$ und $\hat{\pi}_{i'j'}$, $ij \neq i'j'$, gilt:

$$\text{Cov}(\hat{\pi}_{ij}, \hat{\pi}_{i'j'}) = -n\pi_{ij}\pi_{i'j'}.$$

5.2 Gütemaße

Analog zum Kapitel über Bewertungsmaße für Konfidenzintervalle für eine Erfolgswahrscheinlichkeit einer Bernoulliverteilung werden nun Größen zur Bewertung von Konfidenzintervallen für die Differenz zweier verbundener Raten $p_1 - p_2$ betrachtet. Als Abhängigkeitsmaß wird das Odds-Ratio herangezogen.

Definition 5.2.1 (COVERAGED-PROBABILITY)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio¹ sei θ und $\mathcal{KI}(X_1, X_2)$ ein Konfidenzintervall für die verbundene Differenz $p_1 - p_2$. Das Maß

$$P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2))$$

heißt Coverage-Probability des Intervalls \mathcal{KI} .

Die Abbildungen

$$C_{n,\theta} : (0, 1)^2 \rightarrow [0, 1], \text{ wobei } (p_1, p_2) \mapsto P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)), \text{ und} \quad (5.2)$$

$$C_{n,\theta,p_2} : (0, 1) \rightarrow [0, 1], \text{ wobei } p_1 \mapsto P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)), \quad (5.3)$$

heißen Coverage-Prozess für gegebenes Odds-Ratio θ und Coverage-Prozess für gegebenes p_2 und gegebenes θ des Konfidenzintervalls \mathcal{KI} .

Anmerkung 5.2.2 In den folgenden Definitionen wird das Odds-Ratio aus Gründen der Vollständigkeit erwähnt. Der Grund ist, dass die Coverage-Probability in dieser Arbeit aus einem fest vorgegebenem Odds-Ratio berechnet wird (vgl. Definition 5.2.1).

Lemma 5.2.3 (EXAKTE VERBUNDENE COVERAGED-PROBABILITY)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio sei θ und $\mathcal{KI}(X_1, X_2)$ ein Konfidenzintervall für $p_1 - p_2$. Es gilt:

$$\begin{aligned} & P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)) \\ &= P(\pi_{12} - \pi_{21} \in \mathcal{KI}(X_1, X_2)) \\ &= C_{n,\theta}(p_1, p_2) = \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{p_1 - p_2 \in \mathcal{KI}(i,j)} \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot \pi_{21}^j \cdot (1 - \pi_{12} - \pi_{21})^{n-i-j}, \end{aligned}$$

wobei $\mathbb{I}_{(\cdot)}$ eine Indikatorabbildung bezeichnet, mit

$$\mathbb{I}_{p_1 - p_2 \in \mathcal{KI}(i,j)} = \begin{cases} 1 & , \text{ falls } p_1 - p_2 \in \mathcal{KI}(i, j), \\ 0 & , \text{ falls } p_1 - p_2 \notin \mathcal{KI}(i, j). \end{cases}$$

Beweis:

Es gilt in Anlehnung an die Vierfeldertafel:

$$\begin{aligned} & P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)) \\ &= P(\pi_{11} + \pi_{12} - \pi_{11} - \pi_{21} \in \mathcal{KI}(X_1, X_2)) \\ &= E(\mathbb{I}_{\pi_{12} - \pi_{21} \in \mathcal{KI}(X_1, X_2)}) \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{p_1 - p_2 \in \mathcal{KI}(i,j)} \cdot P(X_1 = i, X_2 = j) \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{p_1 - p_2 \in \mathcal{KI}(i,j)} \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot \pi_{21}^j \cdot (1 - \pi_{12} - \pi_{21})^{n-i-j}. \end{aligned}$$

□

¹Die Coverage-Probability wird iterativ aus einem vorgegebenem Odds-Ratio berechnet. Ohne dessen Vorgabe ist die verbundene Coverage-Probability nicht berechenbar. Nähere Details und Angaben sind im Abschnitt 5.3 angegeben.

Lemma 5.2.4 (EIGENSCHAFTEN DER COVERAGE-KURVEN)

Es gelte das Modell (6.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio sei θ und $\mathcal{KI}(X_1, X_2)$ ein Konfidenzintervall für $p_1 - p_2$. Die Abbildung $C_{n, \theta, p_2} : (0, 1) \rightarrow [0, 1]$ sei der Coverage-Prozess für gegebenes p_2 und gegebenes Odds-Ratio des Konfidenzintervalls \mathcal{KI} . Dann gilt:

- (1) $C_{n, \theta, p_2}(p_1)$ ist unstetig.
- (2) Die Abbildung $C_{n, \theta, \frac{1}{2}}(p_1)$ ist symmetrisch um $p_1 = \frac{1}{2}$.
- (3) $P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)) = P(p_2 - p_1 \in \mathcal{KI}(X_1, X_2))$.

Beweis:

- (1) Die Unstetigkeit folgt aus der Existenz des Indikators.
- (2) Es sei ein beliebiges $\delta \in (0, \frac{1}{2})$ gewählt. Nach Modell (5.1) gilt:

$$\begin{aligned} \left(\frac{1}{2} + \delta\right) &= \pi_{11} + \pi_{12} \\ \Leftrightarrow -\delta &= \pi_{21} - \pi_{12} \\ \Leftrightarrow \pi_{21} &= \pi_{12} - \delta. \end{aligned}$$

Damit gilt schließlich:

$$\begin{aligned} &C_{n, \theta, \frac{1}{2}}\left(\frac{1}{2} + \delta\right) \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{\delta \in \mathcal{KI}(i, j)} \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot (\pi_{12} - \delta)^j \cdot (1 - \pi_{12} - (\pi_{12} - \delta))^{n-i-j} \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{\delta \in \mathcal{KI}(i, j)} \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot (-\delta + \pi_{12})^j \cdot (1 - 2\pi_{12} + \delta)^{n-i-j} \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \mathbb{I}_{-\delta \in \mathcal{KI}(i, j)} \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot (\delta + \pi_{12})^j \cdot (1 - 2\pi_{12} - \delta)^{n-i-j} \\ &= C_{n, \theta, \frac{1}{2}}\left(\frac{1}{2} - \delta\right). \end{aligned}$$

$$\begin{aligned} (3) \quad P(p_1 - p_2 \in \mathcal{KI}(X_1, X_2)) &= P(\pi_{12} - \pi_{21} \in \mathcal{KI}(X_1, X_2)) \\ &\stackrel{(2)}{=} P(\pi_{21} - \pi_{12} \in \mathcal{KI}(X_1, X_2)) = P(p_2 - p_1 \in \mathcal{KI}(X_1, X_2)). \end{aligned}$$

□

Aus Lemma 5.2.4 folgt, dass mit $p_2 = \frac{1}{2}$ bedingte Coverage-Kurven stets symmetrisch um $p_1 = \frac{1}{2}$ verlaufen. Die Symmetrie ist unabhängig vom Odds-Ratio. Diese Aussage ist für die Programmierung von wichtiger Bedeutung. Die Symmetrie der bedingten Kurve überträgt sich auf den unbedingten Verlauf. Bei der Programmierung der Maße genügt es daher, nicht den gesamten Definitionsbereich $(0, 1)^2$ zu betrachten, sondern es reicht der Raum $(0, \frac{1}{2})^2$. Selbst mit diesem Zwischenergebnis ist der Programmieraufwand enorm. Man beachte, dass für eine gute Schätzung der Maße aus Definition 5.2.5 mittels numerischer Integration je Integrationsachse mindestens 1000 Stützstellen vorgegeben werden müssen. Besser wären noch 10.000. Gibt man dazu einen Stichprobenumfang von $n=40$ vor, dann müssen $10.000 \times 10.000 \times (40+1) \times (40+1) = 168100000000$

Berechnungen und Indikatorüberprüfungen durchgeführt werden. Ein Stichprobenumfang von $n=40$ ist sogar in verbundenen Designs sehr wenig. Zu beachten ist hier, dass ohne Matrizenprogrammierung derartige Berechnungen nicht durchführbar sind. Das heißt, das Programm muss 10.000×10.000 verschiedene $(40+1) \times (40+1)$ Felder aus Ober- und Untergrenzen bestimmen und in diesen die Indikatoren berechnen. Im Programmpaket 'R' dauert die numerische Berechnung eines derartigen Doppelintegrals mehrere Tage und ist für höhere Stichprobenumfänge nicht mehr durchführbar.

Definition 5.2.5 (MITTLERE VERBUNDENE COVERAGE-PROBABILITY)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio sei θ und $C_{n,\theta}(p_1, p_2)$ ein Coverage-Prozess für gegebenes Odds-Ratio des Intervalls $\mathcal{KI}(X_1, X_2)$. Die Größen

$$\overline{CV}_{n,\theta,p_2} = \int_0^1 C_{n,\theta,p_2}(p_1) dp_1 \quad \text{und} \quad (5.4)$$

$$\overline{CV}_{n,\theta} = \int_0^1 \int_0^1 C_{n,\theta}(p_1, p_2) dp_1 dp_2 \quad (5.5)$$

heißen mittlere Coverage-Probability für gegebenes p_2 und Odds-Ratio θ und mittlere Coverage-Probability für gegebenes θ des Konfidenzintervalls \mathcal{KI} .

Definition 5.2.6 (VERBUNDENE LÄNGENKURVE)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$.

$\mathcal{KI}(X_1, X_2) = [\mathcal{U}(X_1, X_2), O(X_1, X_2)]$ sei ein Konfidenzintervall für $p_1 - p_2$. Das Odds-Ratio² sei θ . Die Abbildungen

$$L_{n,\theta} : (0, 1)^2 \rightarrow [0, 1], \text{ wobei } (p_1, p_2) \mapsto E(O(X_1, X_2) - \mathcal{U}(X_1, X_2)), \text{ und} \quad (5.6)$$

$$L_{n,\theta,p_2} : (0, 1) \rightarrow [0, 1], \text{ wobei } p_1 \mapsto E_{p_2}(O(X_1, X_2) - \mathcal{U}(X_1, X_2)), \quad (5.7)$$

heißen Längenkurve für gegebenes Odds-Ratio θ und Längenkurve für gegebenes p_2 und gegebenes θ des Konfidenzintervalls \mathcal{KI} .

Anmerkung 5.2.7 Die Notation $E_{p_2}(\cdot)$ bedeutet, dass der bedingte Erwartungswert mit gegebenem p_2 betrachtet wird. Die übliche Notation für bedingte Erwartungswerte wird hier aus Gründen der Einheitlichkeit nicht verwendet.

Lemma 5.2.8 (ERWARTETE LÄNGE)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. $\mathcal{KI}(X_1, X_2)$ sei ein Konfidenzintervall für $p_1 - p_2$. Für die erwartete Länge des Konfidenzintervalls gilt:

$$\begin{aligned} & E(O(X_1, X_2) - \mathcal{U}(X_1, X_2)) \\ &= L_{n,\theta}(p_1, p_2) \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} [O(i, j) - \mathcal{U}(i, j)] \cdot \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot \pi_{21}^j \cdot (1 - \pi_{12} - \pi_{21})^{n-i-j}. \end{aligned}$$

Beweis:

Die Behauptung folgt direkt aus der Definition des Konfidenzintervalls $\mathcal{KI}(X_1, X_2)$. \square

²Analog zu Definition 5.2.1 wird die erwartete Länge iterativ aus einem vorgegebenem Odds-Ratio berechnet.

Definition 5.2.9 (MITTLERE LÄNGE)

Es gelte das Modell (5.1). Das Odds-Ratio sei θ und $L_{n,\theta}(p_1, p_2)$ eine Längenkurve. Die Größen

$$\overline{LV}_{n,\theta} = \int_0^1 \int_0^1 L_{n,\theta}(p_1, p_2) dp_1 dp_2 \quad \text{und} \quad (5.8)$$

$$\overline{LV}_{n,\theta,p_2} = \int_0^1 L_{n,\theta,p_2}(p_1) dp_1 \quad (5.9)$$

heißen mittlere Länge für gegebenes Odds-Ratio und mittlere Länge für gegebenes p_2 und Odds-Ratio θ des Konfidenzintervalls \mathcal{KI} .

Definition 5.2.10 (MITTLERE KONSERVATIVITÄT UND ANTI-KONSERVATIVITÄT)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio sei θ und $C_{n,\theta}(p_1, p_2)$ eine Coverage-Kurve für gegebenes Odds-Ratio des Konfidenzintervalls $\mathcal{KI}(X_1, X_2)$. Die Größen

$$\overline{KV}_{n,\theta,p_2} = \int_{\{C_{n,\theta,p_2}(p_1) \geq (1-\alpha)\}} C_{n,\theta,p_2}(p_1) dp_1 \quad \text{und} \quad (5.10)$$

$$\overline{KV}_{n,\theta} = \int \int_{\{C_{n,\theta}(p_1, p_2) \geq (1-\alpha)\}} C_{n,\theta}(p_1, p_2) dp_1 dp_2 \quad \text{bzw.} \quad (5.11)$$

$$\overline{AKV}_{n,\theta,p_2} = \int_{\{C_{n,\theta,p_2}(p_1) \leq (1-\alpha)\}} C_{n,\theta,p_2}(p_1) dp_1 \quad \text{und} \quad (5.12)$$

$$\overline{AKV}_{n,\theta} = \int \int_{\{C_{n,\theta}(p_1, p_2) \leq (1-\alpha)\}} C_{n,\theta}(p_1, p_2) dp_1 dp_2 \quad (5.13)$$

heißen mittlere Konservativität für gegebenes p_2 und gegebenes Odds-Ratio und mittlere Konservativität für gegebenes Odds-Ratio bzw. mittlere Anti-Konservativität für gegebenes p_2 und gegebenes Odds-Ratio und mittlere Anti-Konservativität für gegebenes Odds-Ratio θ des Konfidenzintervalls \mathcal{KI} .

Definition 5.2.11 (MITTLERER ABSOLUTER FEHLER)

Es gelte das Modell (5.1). Das Odds-Ratio sei θ und $C_{n,\theta}(p_1, p_2)$ eine Coverage-Kurve des Konfidenzintervalls $\mathcal{KI}(X_1, X_2)$. Die Größen

$$\overline{F}_{n,\theta,p_2} = \int_0^1 |C_{n,\theta,p_2}(p_1) - (1-\alpha)| dp_1 \quad \text{und} \quad (5.14)$$

$$\overline{F}_{n,\theta} = \int_0^1 \int_0^1 |C_{n,\theta}(p_1, p_2) - (1-\alpha)| dp_1 dp_2 \quad (5.15)$$

heißen mittlerer Absoluter Fehler für gegebenes p_2 und Odds-Ratio θ und mittlerer Fehler für gegebenes Odds-Ratio θ des Konfidenzintervalls \mathcal{KI} .

Definition 5.2.12 (MITTLERE STREUUNG)

Es gelte das Modell (5.1) und es seien $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Das Odds-Ratio sei θ und $C_{n,\theta}(p_1, p_2)$ eine Coverage-Kurve mit gegebenem Odds-Ratio des Konfidenzintervalls $\mathcal{KI}(X_1, X_2)$. Die Größen

$$\overline{S}_{n,\theta,p_2} = \sqrt{\int_0^1 (C_{n,\theta,p_2}(p_1) - (1-\alpha))^2 dp_1} \quad \text{und} \quad (5.16)$$

$$\overline{S}_{n,\theta} = \sqrt{\int_0^1 \int_0^1 (C_{n,\theta}(p_1, p_2) - (1-\alpha))^2 dp_1 dp_2} \quad (5.17)$$

heißen mittlere Streuung für gegebenes p_2 und Odds-Ratio θ und mittlere Streuung für gegebenes Odds-Ratio θ des Konfidenzintervalls $\mathcal{K}I$.

5.3 Konstruktion der verbundenen Coverage-Kurven

Die verbundenen Coverage-Kurven werden iterativ aus einem gegebenem Odds-Ratio berechnet. Hervorgehoben sei hier, dass die Coverage-Probability eine Funktion von p_1 und p_2 ist. Zunächst beobachtet man den Sachverhalt:

$$p_1 = \pi_{11} + \pi_{12}, \quad p_2 = \pi_{11} + \pi_{21}. \quad (5.18)$$

Gibt man sich nun ein Odds-Ratio als fest vor, dann ergibt sich unter der Verwendung von (5.18):

$$\theta = \frac{\pi_{11} \cdot \pi_{22}}{\pi_{21} \cdot \pi_{12}} = \frac{\pi_{11} \cdot (1 - \pi_{11} - \pi_{12} - \pi_{21})}{(p_1 - \pi_{11}) \cdot (p_2 - \pi_{11})} = \frac{\pi_{11} \cdot (1 - p_1 - p_2 + \pi_{11})}{(p_1 + \pi_{11}) \cdot (p_2 - \pi_{11})}.$$

Damit gilt für $\theta = 1$:

$$\pi_{11} = p_1 \cdot p_2.$$

Setzt man $\theta \neq 1$, dann gilt:

$$\pi_{11} = \underbrace{\frac{(\theta - 1) \cdot p_2 + (\theta - 1) \cdot p_1 + 1}{2 \cdot (\theta - 1)}}_A \begin{cases} -\sqrt{A^2 - \frac{\theta \cdot p_2 \cdot p_1}{(\theta - 1)}} & \theta > 1, \\ +\sqrt{A^2 - \frac{\theta \cdot p_2 \cdot p_1}{(\theta - 1)}} & 0 \leq \theta < 1. \end{cases}$$

Mit diesem Parameter werden nun die Parameter π_{12} und π_{21} mit (5.18) berechnet.

5.4 Konfidenzintervalle

Als Grundlage zur Herleitung der Konfidenzintervalle dient das verbundene Modell (5.1). Die Erfolgswahrscheinlichkeiten p_1 und p_2 werden erwartungstreu durch die relativen Häufigkeiten gemäß (4.1) geschätzt.

Verwendet man relative Häufigkeiten als erwartungstreu Schätzer für die Erfolgswahrscheinlichkeiten, dann gilt $E(\hat{p}_1 - \hat{p}_2) = E(\hat{\pi}_{12} - \hat{\pi}_{21})$. Weiterhin gilt:

$$\text{Var}(\hat{\pi}_{12} - \hat{\pi}_{21}) = \text{Var}(\hat{\pi}_{12}) + \text{Var}(\hat{\pi}_{21}) - 2\text{Cov}(\hat{\pi}_{12}, \hat{\pi}_{21}) \quad (5.19)$$

$$= \frac{1}{n} \left(\pi_{12}(1 - \pi_{12}) + \pi_{21}(1 - \pi_{21}) + 2\pi_{12}\pi_{21} \right). \quad (5.20)$$

Die Varianz zerlegt sich somit in die Summe der Varianzen der relativen Häufigkeiten $\hat{\pi}_{12}$ und $\hat{\pi}_{21}$. Einen konsistenten Schätzer für (5.19) liefert das folgende Lemma.

Lemma 5.4.1 (KONSISTENTER SCHÄTZER FÜR $\text{Var}(\hat{p}_1 - \hat{p}_2)$)

Es gelte das verbundene Modell (5.1). Unter der Verwendung von (5.19) bezeichnet

$$\widehat{\text{Var}}(\hat{p}_\Delta) = \widehat{\text{Var}}(\hat{\pi}_{12} - \hat{\pi}_{21}) = \frac{1}{n} \left(\hat{\pi}_{12}(1 - \hat{\pi}_{12}) + \hat{\pi}_{21}(1 - \hat{\pi}_{21}) + 2\hat{\pi}_{12}\hat{\pi}_{21} \right)$$

einen konsistenten Schätzer für $\text{Var}(\hat{\pi}_{12} - \hat{\pi}_{21})$.

Beweis:

Die Summe von konsistenten Schätzern ist wieder konsistent. Die Behauptung folgt schließlich durch die Verwendung von (4.18). \square

Mit diesen Ergebnissen sind Statistiken aufstellbar. Als einfachste unter diesen ist die Wald-Statistik (vgl. 4.21) für verbundene Stichproben bekannt. Die Invertierung dieser liefert das sogenannte 'Wald'-Konfidenzintervall für verbundene Stichproben, dessen Grenzen durch

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2} \cdot \sqrt{[(\hat{\pi}_{12} + \hat{\pi}_{21}) - (\hat{\pi}_{21} - \hat{\pi}_{12})^2]/n} \quad (5.21)$$

gegeben sind. Betrachtet man (5.21), dann fallen drei Nachteile dieses Intervalls auf:

- Falls $\hat{\pi}_{12} = \hat{\pi}_{21} = 0$ ist, dann folgt für die Grenzen des Intervalls: $\mathcal{KI} = [\mathcal{U}, \mathcal{O}] = [0, 0]$. Im Allgemeinen gilt jedoch, dass $\pi_{12} > 0$ und $\pi_{21} > 0$ sind, sodass der wahre Standardfehler positiv ist.
- Falls $\hat{p}_1 = 0$ ist, dann vereinfacht sich das Intervall zum Wald-Intervall für $-p_2$ und übernimmt damit die schlechten Eigenschaften des Wald-Intervalls für eine Stichprobe.
- Das Zentrum des Intervalls liegt in $\hat{p}_1 - \hat{p}_2$. Analog zum Einstichprobenproblem ist eine derartige Schätzung der Differenz nicht empfehlenswert.

Die äußerst geringe Coverage-Probability des Wald-Intervalls für einen Parameter überträgt sich auf das Wald-Konfidenzintervall für die Differenz zweier verbundener Raten. Eine Veränderung des Stichprobenumfangs um zwei Erfolge und zwei Misserfolge veränderte den Punktschätzer für die unbekannte Erfolgswahrscheinlichkeit und verbesserte deutlich die Coverage-Probability des Wald-Konfidenzintervalls. Überträgt man diese Idee auf verbundene Stichproben, dann bedeutet dieses, dass nun alle Erfolgswahrscheinlichkeitsschätzer für die Parameter der Vierfeldertafel verändert werden müssen. In der Arbeit von AGRESTI und MIN (2005) [20] wurde das Verhalten von derartig modifizierten Wald-Konfidenzintervallen untersucht. Verändert man also den Stichprobenumfang n durch

$$n^* = n + t, \quad t \in \mathbb{N},$$

dann folgt daraus, dass die Schätzer für p_1 bzw. p_2 folgendermaßen modifiziert werden:

$$\hat{p}_1^* \rightarrow \frac{n_{11} + n_{12} + t/2}{n + t}, \quad \hat{p}_2^* \rightarrow \frac{n_{11} + n_{21} + t/2}{n + t}.$$

Setzt man diese schließlich in die Formel (5.21) ein, dann ergibt sich das Agresti-Intervall für die verbundene Differenz $p_1 - p_2$ durch

$$\frac{n_{21}^* - n_{12}^*}{n^*} \pm z_{1-\alpha/2} \cdot \frac{\sqrt{(n_{12}^* + n_{21}^*) - [(n_{12} - n_{21}^*)^2/n^*]}}{n^*}, \quad (5.22)$$

$$\text{wobei} \quad n_{12}^* = n_{12} + t/4, \quad n_{21}^* = n_{21} + t/4, \quad n^* = n + t,$$

gilt.

Der große Nachteil dieses Intervalls liegt darin, dass es nicht bereichserhaltend ist. Bei Grenzüberschreitung werden die Überstände abgeschnitten und auf +1 bzw. -1 gesetzt.

AGRESTI und MIN (2005) [20] untersuchten für $t = 1, 2, 3, 4$ das Verhalten der Wald+ t -Intervalle und kamen zu dem Ergebnis, dass für $t = 2$, also $+1/2$ für jede Zelle, die Konfidenzintervalle eine gute Coverage-Probability aufweisen und kaum vom 95% KONfidenzniveau abweichen (vgl.

[20]). Dazu dominiert der Aspekt, dass diese sehr leicht zu berechnen sind und damit auch in elementare Kurse über Statistik als Vorlesungsstoff integriert werden können.

Hervorgehoben sei hier aber der Fakt, dass das Vorgehen einem heuristischen Verfahren folgt und damit keine analytische Begründung dieses Intervalls möglich ist. Ein derartiges Vorgehen ist damit nicht auf mehrdimensionale Versuchsanlagen übertragbar.

Ein Resultat aus dem Einstichprobenproblem war, dass das Score-Intervall von Wilson äußerst gute Eigenschaften aufweist. Überraschenderweise wurde erst in der Arbeit von TANGO (1998) [26] ein Score-Test invertiert. Das Score-Intervall beinhaltet alle Parameter

$$\Delta = p_1 - p_2,$$

für die gilt:

$$\frac{|(\hat{p}_1 - \hat{p}_2) - \Delta|}{\sqrt{[(\hat{\pi}_{12}(\Delta) + \hat{\pi}_{21}(\Delta)) - \Delta^2]/n}} < z_{\alpha/2}.$$

Die Angabe $\hat{\pi}_{ij}(\Delta)$ bezeichnet dabei den Maximum-Likelihood Schätzer von π_{ij} unter der Hypothese $H_0 : (p_1 - p_2) = \Delta$ mit

$$\begin{aligned}\hat{\pi}_{21}(\Delta) &= (\sqrt{B^2 - 4AC} - B)/(2A), \\ A &= 2n, \\ B &= -b - c + (2n - b + c)\Delta, \\ C &= -c\Delta(1 - \Delta).\end{aligned}$$

In der Literatur gibt es keine einheitliche Lösung für dieses Konfidenzintervall. Man erhält die Grenzen dieses Intervalls mit einem iterativen Verfahren, welches ausführlich in den Arbeiten von TANG, TANG, CHAN (2005) [28] und NEWCOMBE (1998) [14] beschrieben wird.

In der Literatur sind zahlreiche andere, meist sehr anspruchsvolle, Konfidenzintervalle bekannt. In dieser Arbeit wurden die aus den aktuellsten Papern [20] und [28] mit der besten Bewertung ausgewählt.

5.4.1 Programmierung des Score-Intervalls

Das asymptotische Score-Intervall für die Differenz zweier verbundener Raten wurde iterativ mit Hilfe eines Suchalgorithmus berechnet. Der Suchalgorithmus umfasst dabei die Werte von Δ , die der obigen Bedingung genügen. Der Algorithmus zur Berechnung lautet folgendermaßen:

```
scoreci<-function(b,c,n,conflev)  #\Start mit Untergrenze
{  pa = 2*n                       #\ Übergebe A
  z = qnorm(1-(1-conflev)/2)     # \Konfidenzlevel
  if(c == n) {ul = 1}           # \Vollergebnis
  else{
    proot = (c-b)/n
    dp = 1-proot
    niter = 1
    while(niter <= 50){         #\starte das suchverfahren
      dp = 0.5*dp
      up2 = proot+dp           # \ definiere die untergrenze
      pb = - b - c + (2*n-c+b)*up2  # \B
      pc = -b*up2*(1-up2)      #\C
      q21 = (sqrt(pb^2-4*pa*pc)-pb)/(2*pa) #\ML-Schätzer
      score = (c-b-n*up2)/sqrt(n*(2*q21+up2*(1-up2))) #\ Teststatistik
      if(abs(score)<z){ proot = up2 }
      niter=niter+1
      if((dp<0.0000001) || (abs(z-score)<.000001)){
```

```

niter=51
ul=up2
}
}
if(b == n) {ll = -1}          #\Obergernze
else{
  proot = (c-b)/n
  dp = 1+proot
  niter = 1
  while(niter <= 50){
    dp = 0.5*dp
    low2 = proot-dp
    pb = - b - c + (2*n-c+b)*low2
    pc = -b*low2*(1-low2)
    q21 = (sqrt(pb^2-4*pa*pc)-pb)/(2*pa)
    score = (c-b-n*low2)/sqrt(n*(2*q21+low2*(1-low2)))
    if(abs(score) < z){proot = low2}
    niter = niter+1
    if((dp<0.0000001) || (abs(z-score)<.000001)){
      ll = low2
      niter = 51
    }
  }
}
c(ll,ul)
}

```

5.4.2 Transformierte Konfidenzintervalle

Das Ziel ist, ein Konfidenzintervall für die Differenz $p_\Delta = p_1 - p_2$ zu erhalten. Dieses bedeutet, dass unter der Verwendung der Transformationsmethode die Differenz stetig transformiert werden muss. Die Funktion muss natürlich den Anforderungen aus der Transformationsmethode (siehe Kapitel 3.4) genügen.

Definition 5.4.2 (FOLGE)

Es gelte $n \in \mathbb{N} - \{0\}$. Die Abbildung

$$c : \mathbb{N} - \{0\} \rightarrow \mathbb{R}, \quad n \mapsto c(n),$$

heißt Folge. Die Abbildung $c : \mathbb{N} - \{0\} \rightarrow \mathbb{R}$ heißt Nullfolge, falls $\lim_{n \rightarrow \infty} c(n) = 0$ gilt.

Aus Definition 5.4.2 folgt, dass sich insbesondere Nullfolgen leicht konstruieren lassen. Es gelte $v \in \mathbb{Q}_{\geq 0}$ und $\lambda \in \mathbb{N} - \{0\}$. Dann ist für $v < n$, wobei v fest sei, durch

$$c : \mathbb{N} - \{0\} \rightarrow \mathbb{R}, \quad n \mapsto \frac{v}{n^\lambda}, \tag{5.23}$$

eine Nullfolge definiert.

Satz 5.4.3 (TRANSFORMIERTE VERBUNDENE STATISTIK)

Es gelte das Modell (5.1). Die Funktion $g : (-1, 1) \rightarrow \mathbb{R}$ erfülle die Anforderungen an eine Transformationsfunktion aus Kapitel 3.4. Die Differenz der relativen Häufigkeiten sei $\hat{p}_\Delta = \frac{1}{n} \sum_{k=1}^n (X_{1k} - X_{2k})$. Durch die Verwendung des konsistenten Schätzers aus Lemma (5.4.1) für $\text{Var}(\hat{p}_\Delta)$ gilt:

$$\sqrt{n} \left(\frac{g(\hat{p}_\Delta) - g(p_\Delta)}{\left(\frac{dg(\hat{p}_\Delta)}{d\hat{p}_\Delta} \right) \sqrt{\widehat{\text{Var}}(\hat{p}_\Delta)}} \right) \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1). \tag{5.24}$$

Inbesondere gilt, dass durch den Einbau einer Nullfolge in den Varianzschätzer gemäß (5.23) durch

$$\text{Var}(\widehat{\hat{p}_1 - \hat{p}_2}) \hookrightarrow \text{Var}(\widehat{\hat{p}_1 - \hat{p}_2}) + c(n), \quad c(n) = \frac{v}{n^k}, \quad (5.25)$$

die Aussage (5.24) gültig bleibt.

Beweis:

Da $X_{jk} \stackrel{u.i.v.}{\sim} \mathcal{B}(p_j)$ gilt, gilt auch:

$$E(X_{11} - X_{21}) = p_1 - p_2 = p_\Delta, \quad \text{Var}(X_{11} - X_{21}) = n \cdot \text{Var}(\hat{p}_\Delta).$$

Die Behauptung folgt schließlich durch die Anwendung der δ -Methode aus Satz A.0.8, des Satzes von Slutsky A.0.9, des zentralen Grenzwertsatzes A.0.6 und Lemma 5.4.1. \square

Korollar 5.4.4 (TRANSFORMIERTE KONFIDENZINTERVALLE FÜR VERBUNDENE LINEAR-KOMBINATIONEN)

Es gelten die Voraussetzungen aus Satz 5.4.3. Ein asymptotisches und bereichserhaltendes $(1 - \alpha)$ -Konfidenzintervall für p_Δ ist gegeben durch:

$$[\mathcal{U}, \mathcal{O}] = [g^{-1}(\mathcal{U}^{Transf}), g^{-1}(\mathcal{O}^{Transf})] \subset (0, 1),$$

wobei

$$[\mathcal{U}^{Transf}, \mathcal{O}^{Transf}] = g(\hat{p}_\Delta) \mp \frac{z_{1-\alpha/2}}{\sqrt{n}} \left(\frac{dg(\hat{p}_\Delta)}{d\hat{p}_\Delta} \right) \sqrt{\text{Var}(\widehat{\hat{p}_1 - \hat{p}_2})} \quad (5.26)$$

gilt.

Beweis:

Wegen Satz 5.4.3 und aufgrund der Monotonie und Umkehrbarkeit von g folgt:

$$\begin{aligned} P(p_\Delta \in [\mathcal{U}, \mathcal{O}]) &= P(g(p_\Delta) \in [[\mathcal{U}^{Transf}, \mathcal{O}^{Transf}]]) \\ &= P\left(z_{\alpha/2} \leq \sqrt{n} \left(\frac{g(\hat{p}_\Delta) - g(p_\Delta)}{\left(\frac{dg(\hat{p}_\Delta)}{d\hat{p}_\Delta} \right) \sqrt{\text{Var}(\widehat{\hat{p}_1 - \hat{p}_2})}} \right) \leq z_{1-\alpha/2} \right) \xrightarrow{n \rightarrow \infty} 1 - \alpha. \end{aligned}$$

Die Behauptung folgt durch die Anwendung der Gleichung: $z_{\alpha/2} = -z_{1-\alpha/2}$. \square

Mit diesen Ergebnissen können nun explizit Konfidenzintervalle durch die Betrachtung von Transformationsfunktionen berechnet werden. Die Funktion muss natürlich die Anforderungen an eine Transformationsfunktion ((T1) – (T6)) erfüllen.

Lemma 5.4.5 (TRANSFORMIERTE INTERVALLE)

Es gelten die Voraussetzungen von Satz 5.4.3. Es sei $c(n) = \frac{v}{n^k}$ eine Nullfolge gemäß (5.23). Durch die Betrachtung und Anwendung der Funktionenscharen

$$g_k : (-1, 1) \rightarrow \mathbb{R}, \quad (p_1 - p_2) = p_\Delta \mapsto \frac{1}{2} \cdot \log \left(\frac{1+k+p_\Delta}{1+k-p_\Delta} \right), \quad (5.27)$$

mit den Ableitungen

$$\frac{dg_k(p_\Delta)}{dp_\Delta} = \frac{1+k}{(1+k)^2 - p_\Delta^2} \quad (5.28)$$

und den Umkehrfunktionen

$$g^{-1} : \mathbb{R} \rightarrow (0, 1), \quad \text{wobei } x \mapsto \frac{\exp(2x) - 1}{\exp(2x) + 1}, \quad (5.29)$$

dann sind durch

$$P(g^{-1}(\mathcal{U}^T) \leq p_1 - p_2 \leq g^{-1}(\mathcal{O}^T)) \approx 1 - \alpha, \quad (5.30)$$

wobei

$$\mathcal{U}^T, \mathcal{O}^T = g_k(\hat{p}_\Delta) \pm z_{1-\alpha/2} \cdot \frac{dg_k(\hat{p}_\Delta)}{d\hat{p}_\Delta} \cdot \sqrt{\widehat{\text{Var}}(\hat{p}_1 - \hat{p}_2) + \frac{v}{n\lambda}}$$

gilt, asymptotische und bereichserhaltende $(1 - \alpha)$ -Konfidenzintervalle für p_Δ gegeben.

Beweis:

Die Funktionenscharen erfüllen die Anforderungen an eine Transformationsfunktion. Die Behauptung folgt unmittelbar aus Korollar 5.4.4. \square

Analog zum Einstichprobenproblem wurden in diesem Kapitel Scharen von Konfidenzintervallen hergeleitet. Der Vorteil der Verwendung von derartigen Konfidenzscharen liegt darin, dass keine Null- oder Vollergebnisse abgefangen werden brauchen. Die Motivation zum Einbau einer Nullfolge in den Varianzschätzer liegt darin begründet, dass das neue transformierte Intervall ohne Nullfolge leicht anti-konservatives Verhalten zeigt. Diese schlechte Eigenschaft kann durch den Einbau einer Nullfolge eliminiert werden. Die Konstruktion der Nullfolge ist dadurch motiviert, diese so variabel wie möglich zu gestalten.

Um schließlich ein bestimmtes Intervall aus dieser Schar auswählen zu können, muss eine analytische Begründung für diese Auswahl geliefert werden. Dieses ist schwierig, da die Coverage-Prozesse un stetige Abbildungen darstellen und sich damit mit analytischen Methoden nur sehr schwierig oder eben auch gar nicht analytisch berechnen lassen. Allerdings ist die Längenkurve keine un stetige Abbildung und somit ist die mittlere Länge auch mit mathematischen Methoden berechenbar. Im folgenden Satz wird daher das Verhalten der mittleren Länge der transformierten Intervalle in Abhängigkeit der drei Parametrisierungsparameter untersucht.

Satz 5.4.6 (MONOTONIE DER MITTLEREN LÄNGENKURVE)

Es gelte das Modell (5.1) und $X_j \sim \text{BIN}(n, p_j)$, $j = 1, 2$. Es sei $c(n) = \frac{v}{n\lambda}$ eine Nullfolge gemäß (5.23) und $k \in (0, 1)$ die Parametrisierungskonstante der Fisher-Funktion. $\mathcal{K}I_{k,v,\lambda}(X_1, X_2)$ sei das transformierte Konfidenzintervall aus Lemma 5.4.5 in Abhängigkeit der Parameter (k, v, λ) . $L_{n,\theta,k,v,\lambda}(p_1, p_2)$ sei die zugehörige Längenkurve. Dann sind durch

$$\begin{aligned} (a) \quad \Phi : (0, 1) &\rightarrow [0, 1], \quad k \mapsto \int_0^1 \int_0^1 L_{n,\theta,k,v,\lambda}(p_1, p_2) dp_1 dp_2, \\ (b) \quad \Gamma : \mathcal{D} &\rightarrow [0, 1], \quad v \mapsto \int_0^1 \int_0^1 L_{n,\theta,k,v,\lambda}(p_1, p_2) dp_1 dp_2, \quad \max(\mathcal{D}) < n, \\ (c) \quad \Psi : \mathbb{N} &\rightarrow [0, 1], \quad \lambda \mapsto \int_0^1 \int_0^1 L_{n,\theta,k,v,\lambda}(p_1, p_2) dp_1 dp_2, \end{aligned}$$

streng monoton fallende ((a) und (c)) und durch (b) eine streng monoton steigende Funktion definiert.

Anmerkung 5.4.7 Die Notation im Satz 5.4.6 scheint auf den ersten Blick durch die vielen Indizes etwas verwirrend. Es soll die in dieser Arbeit gewohnte Notation, bedingte Erwartungswerte durch Indizes zu kennzeichnen, nicht geändert werden. Die drei Abbildungen im Satz 5.4.6 beschreiben, wie sich die mittlere Länge des parametrisierten transformierten Intervalls verhält, wenn einer der Parameter verändert wird und die anderen konstant gehalten werden.

Beweis:

Der Beweis verläuft nahezu analog zum Beweis des Satzes 4.3.7. Die Funktionen Φ, Γ und Ψ sind differenzierbar. Weiter ist die Transformationsfunktion bijektiv und streng monoton steigend. Damit kann das Intervall im transformierten System betrachtet werden. Es genügt den Summanden³ $\mathcal{K}I_{k,v,\lambda}(i, j)$ zu betrachten. Es kann auch angenommen werden, dass $(i, j) \neq (0, 0)$. Da weiterhin $P(X_1 = i, X_2 = j) = \frac{n!}{i!j!(n-i-j)!} \cdot \pi_{12}^i \cdot \pi_{21}^j \cdot (1 - \pi_{12} - \pi_{21})^{n-i-j} > 0$ gilt, folgt:

$$\begin{aligned} & [O_{k,v,\lambda}(i, j) - \mathcal{U}_{k,v,\lambda}(i, j)] \\ &= g_k(\hat{p}_1 - \hat{p}_2) + z_{1-\alpha/2} \cdot \frac{(1+k)}{((1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2)} \cdot \sqrt{\left[(\hat{\pi}_{12} + \hat{\pi}_{21}) - (\hat{\pi}_{21} - \hat{\pi}_{12})^2 + \frac{v}{n^\lambda} \right] / n} \\ &- \left[g_k(\hat{p}_1 - \hat{p}_2) - z_{1-\alpha/2} \cdot \frac{(1+k)}{((1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2)} \cdot \sqrt{\left[(\hat{\pi}_{12} + \hat{\pi}_{21}) - (\hat{\pi}_{21} - \hat{\pi}_{12})^2 + \frac{v}{n^\lambda} \right] / n} \right] \\ &= \underbrace{\frac{2 \cdot z_{1-\alpha/2}}{\sqrt{n}}}_{=:K} \underbrace{\frac{(1+k)}{((1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2)}}_{AB} \cdot \underbrace{\sqrt{\left[(\hat{\pi}_{12} + \hat{\pi}_{21}) - (\hat{\pi}_{21} - \hat{\pi}_{12})^2 + \frac{v}{n^\lambda} \right]}}_{=:V}. \end{aligned}$$

Man kann diesen Ausdruck ebenfalls als differenzierbare Funktion auffassen. Für die partiellen Ableitungen gilt:

$$\begin{aligned} (a) \quad \frac{\partial(\cdot)}{\partial k} &= K \cdot V \cdot \left[\frac{-(1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2}{((1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2)^2} \right] \\ &= -K \cdot V \cdot \underbrace{\left[\frac{(1+k)^2 + (\hat{p}_1 - \hat{p}_2)^2}{((1+k)^2 - (\hat{p}_1 - \hat{p}_2)^2)^2} \right]}_{>0, \text{ da Transformationsfunktion}} < 0, \\ (b) \quad \frac{\partial(\cdot)}{\partial v} &= K \cdot AB \cdot \frac{1}{2V} \cdot \frac{1}{n^\lambda} > 0, \\ (c) \quad \frac{\partial(\cdot)}{\partial \lambda} &= -K \cdot AB \cdot \frac{1}{2V} \cdot \frac{v}{n^\lambda} \cdot \ln(n) < 0. \end{aligned}$$

□

Hervorgehoben sei hier, dass die Monotonie unabhängig von der Abhängigkeitsstruktur ist und damit ist das Verfahren unmittelbar auf unabhängige Stichproben übertragbar.

Weiter beobachtet man intuitiv:

- Je größer die Länge, desto höher ist die erwartete Coverage-Probability.
- Die Konservativität steigt mit fallendem k , die Anti-Konservativität steigt mit steigendem k .
- Die Streuung der Coverage-Kurve um das Nominalniveau ist für großes und für kleines k groß.

Resultat 5.4.8 (OPTIMALE PARAMETERKOMBINATION)

Es gelte das Modell (5.1) und die Voraussetzungen von Satz 5.4.6. $\mathcal{K}I_{k,v,\lambda}(X_1, X_2)$ sei das transformierte und parametrisierte Konfidenzintervall in Abhängigkeit der Transformationskonstante k und eingebauter Nullfolge $c(n) = \frac{v}{n^\lambda}$. Die Größen $\bar{S}_{n,\theta,k,v,\lambda}$ bzw. $\bar{F}_{n,\theta,k,v,\lambda}$ seien die mittlere Streuung⁴ bzw. der mittlere Fehler in Abhängigkeit der Parameter des Intervalls $\mathcal{K}I_{k,v,\lambda}(X_1, X_2)$ aus

³Der Summand bezeichnet einen Summanden aus der erwarteten Länge des $\mathcal{K}I_{k,v,\lambda}(X_1, X_2)$ -Intervalls.

⁴Die Aufnahme der Parametrisierungsparameter in den Index bedeutet, dass auf diesen Größen bedingt wird.

den Definitionen 5.2.12 bzw. 5.2.11. Im Mittel führt die Kombination

$$(k^*, v^*, \lambda^*) = \left(\frac{1}{50}, \frac{1}{3}, 1 \right) \quad (5.31)$$

zu einer lokalen minimalen Schwankung des zugehörigen Coverage-Prozesses $C_{n,\theta,k,v,\lambda}(p_1, p_2)$ um die $(1 - \alpha)$ -Ebene.

Notation:

Um den Optimierungsalgorithmus leichter beschreiben zu können, wird nun die mittlere Streuung bzw. der mittlere Fehler der transformierten Konfidenzscharen als Funktion der Parameter angesehen. D.h., es gilt:

$$\begin{aligned} \bar{S}_{n,\theta,k,v,\lambda} &= \bar{S}(n, \theta, k, v, \lambda), \\ \bar{F}_{n,\theta,k,v,\lambda} &= \bar{F}(n, \theta, k, v, \lambda). \end{aligned}$$

Anmerkung 5.4.9 Durch diesen Notationswechsel wird der Algorithmus übersichtlicher. Außerdem wird aus dieser Schreibweise ersichtlich, dass die mittlere Streuung bzw. der mittlere Fehler nicht als Ergebnis betrachtet werden, sondern als formale Grundlage für die Optimierung und damit zur Herleitung des neuen transformierten Intervalls.

Herleitung:

Durch die eingebaute Nullfolge in den Varianzschätzer soll die optimale Parameterkonstellation der Nullfolge sowie die Konstante k bestimmt werden, die zu einer minimalen Schwankung der Coverage-Kurve um die Nominalebene führen. Um ein allgemeines Ergebnis zu erhalten, muss der Optimierungsalgorithmus verschiedene Stichprobenumfänge und Odds-Ratios enthalten.

Der Dreidimensionale Ergebnisraum fordert einen Trilevel Optimierungsalgorithmus aus dem Bereich der multikriteriellen Optimierung. In dieser geht es allgemein darum, das Optimum einer Zielfunktion $Z(x,y,z)$ in x zu finden, unter der Nebenbedingung, dass sowohl y als auch z optimal sind. Vereinfacht wird dieser Sachverhalt, indem der Trilevel Algorithmus auf eine Bilevel Optimierung reduziert wird. Dieses ist möglich, indem zwei der Parameter x,y oder z als Paar angesehen werden. Hier z.B. (x,y) . Die zu optimierende Zielfunktion lautet also in allgemeiner Form:

$$\begin{aligned} Z(x, y, z^*) &\rightarrow \min_{x,y}, \\ &\text{unter den Nebenbedingungen:} \\ z^* &= \min_z Z(x, y, z), \\ (x, y, z) &> 0 \quad (\text{Nichtnegativitätsbedingungen}). \end{aligned}$$

Überträgt man dieses allgemeine Programm, dann gilt es folgendes Optimierungstableau zu lösen:

$$Z_1 = \frac{1}{n_S \cdot n_T} \sum_{j=1}^{n_T} \sum_{i=1}^{n_S} \bar{S}(n_i, \theta_j, k^*, v, \lambda) \rightarrow \min_{(v, \lambda)}, \quad (5.32)$$

unter der Nebenbedingung:

$$k^* = \min_k \frac{1}{n_S \cdot n_T} \sum_{j=1}^{n_T} \sum_{i=1}^{n_S} \bar{S}(n_i, \theta_j, k, v, \lambda), \quad (5.33)$$

$$Z_2 = \frac{1}{n_S \cdot n_T} \sum_{j=1}^{n_T} \sum_{i=1}^{n_S} \bar{F}(n_i, \theta_j, k^*, v, \lambda) \rightarrow \min_{(v, \lambda)}, \quad (5.34)$$

unter der Nebenbedingung:

$$k^* = \min_k \frac{1}{n_S \cdot n_T} \sum_{j=1}^{n_T} \sum_{i=1}^{n_S} \bar{F}(n_i, \theta_j, k, v, \lambda), \quad (5.35)$$

$$(v, \lambda, k) > 0.$$

Die Größen n_S bzw. n_T bezeichnen dabei die Anzahl der Stichprobenumfänge bzw. Odds-Ratios. Das Optimierungsprogramm ist also vereinfacht so zu verstehen, dass das Minimum einer Funktion, welche auf den Mittelwert bestimmter Doppelintegralfunktionen abbildet, in drei Komponenten gesucht ist.

Eindeutigkeit der Lösung:

Aus der Struktur des Programms folgt direkt, dass es keine eindeutige Lösung geben kann, sondern eine bestimmte Menge von Kombinationen, die die bestmöglichen Lösungen enthält. Das Ziel des Algorithmus soll nicht die Lösung an sich, sondern die Qualität der Lösung sein. Notwendigerweise müssen daher vorab Abbruchbedingungen vereinbart werden. Als Lösungsmenge aller Kombinationen $\{v, \lambda, k\}$ soll die Menge angesehen werden, bei welcher sich die Ergebnisse um nicht mehr als $1 * 10^{-5}$ unterscheiden.

Kriterien zur Vorauswahl:

Unter den Kriterien zur Vorauswahl versteht man Verfahren, mit denen bestimmt wird, welche notwendigen Variablen als fest übergeben werden. Wie oben erwähnt, müssen verschiedene Stichprobenumfänge und Odds-Ratios übergeben werden. Ausserdem muss mit einer Paramtermenge mit verschiedenen (v, λ) Konstellationen angefangen werden. Weiterhin sind die Größen n_S und n_T festzulegen.

Als Hauptkriterium zur Vorauswahl dienen in diesem Verfahren die Coverage-Kurven, welche durch sehr viele verschiedene Parameterkonstellationen optisch ausgewertet wurden. Anhand deren wurde entschieden, dass folgende Werte als fest vorgegeben werden:

- Stichprobenumfänge

i= 1	2	3	4	5	6	7
$n_i= 20$	30	40	50	60	70	100.

Diese Auswahl ergab sich, weil sich ab einem Stichprobenumfang von $n = 20$ alle untersuchten Verfahren (Agresti und Transformiert) als anwendbare Verfahren herausstellen. Bei einem Stichprobenumfang von weniger als 20 sind alle Verfahren nicht verwendbar. Begonnen wurde daher mit einem Stichprobenumfang von $n = 20$. Aufgrund der hohen Rechenzeit bei höheren Stichprobenumfängen wurden diese in zehner Schritten erhöht. Ausserdem wurde, um auch ein Resultat für sehr hohe Stichprobenumfänge zu integrieren, ein Stichprobenumfang von $n=100$ implementiert.

- Odds-Ratios

j=	1	2	3	4
$\theta_j=$	1	2	3	4.

Vorgehen:

Der Algorithmus verlangt ein spezielles Suchverfahren. Als Startwert diene eine beliebige Transformationskonstante k mit dem Ziel, im ersten Iterationsschritt die optimale (ν, λ) -Kombination für dieses beliebig vorgegebene k zu finden. Anschließend wurde die optimale (ν, λ) -Kombination in die k -Optimierung eingesetzt. Dieses wurde so lange wiederholt, bis Eindeutigkeit zwischen den Ergebnissen bestand.

Durchführung:

Als numerische Integrationsmethode wurde die zusammengesetzte Trapez-Regel angewendet. Integriert wurde auf 10000 Integrationsstützstellen. Die Konstante k der Transformationsfunktion wurde zum Beginn des Optimierungsalgorithmus durch einen kleinen Startwert ersetzt, so dass man sich beim erstmaligen Vorliegen der (ν, λ) -Kombination zu einem größeren oder noch kleineren k vorarbeiten konnte. Als Startwert diene $k = 1/1000$.

Ergebnis:

Die Ergebnisse der Optimierung werden grafisch in den Abbildungen 5.1 und 5.2 gezeigt.

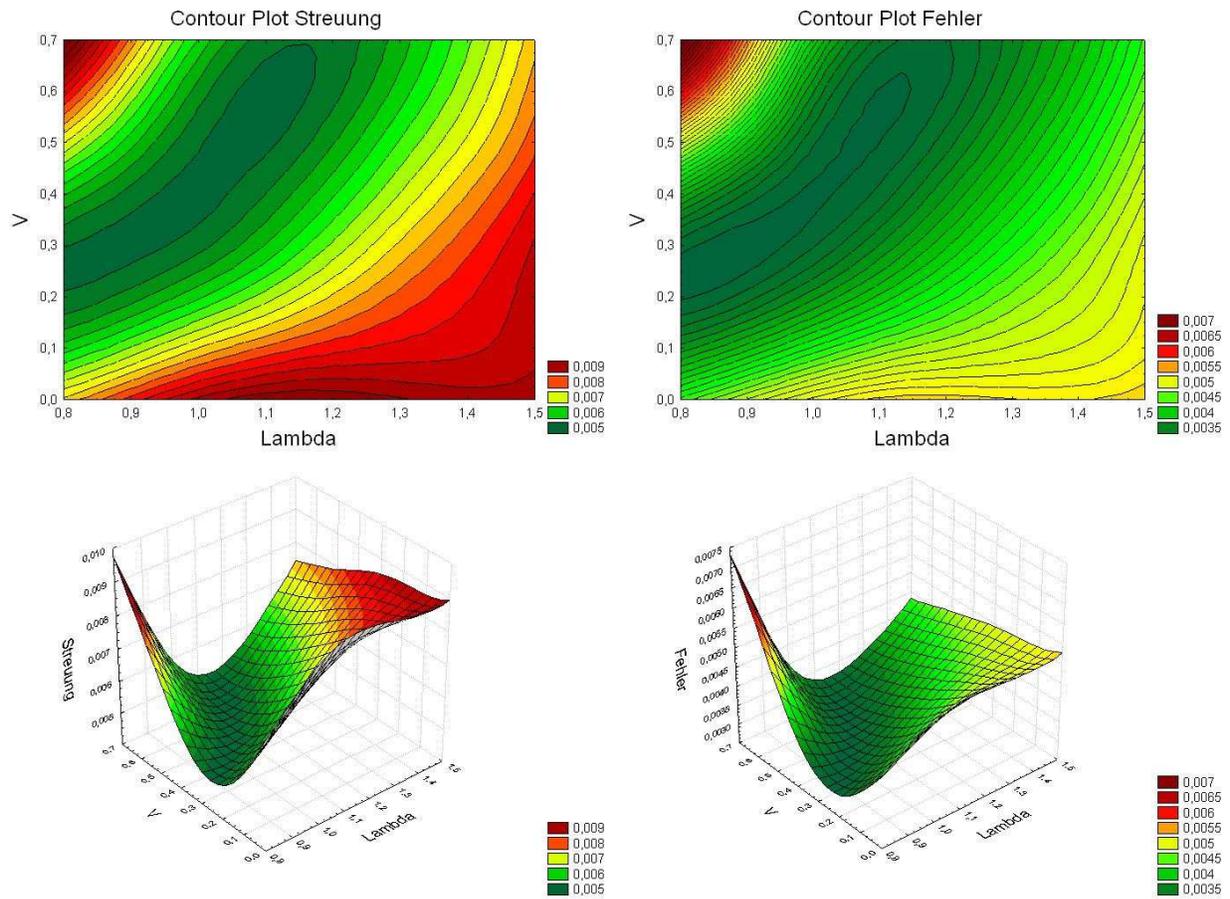


Abbildung 5.1: Contour Plots des zweiten Iterationsschrittes mit $k=1/50$

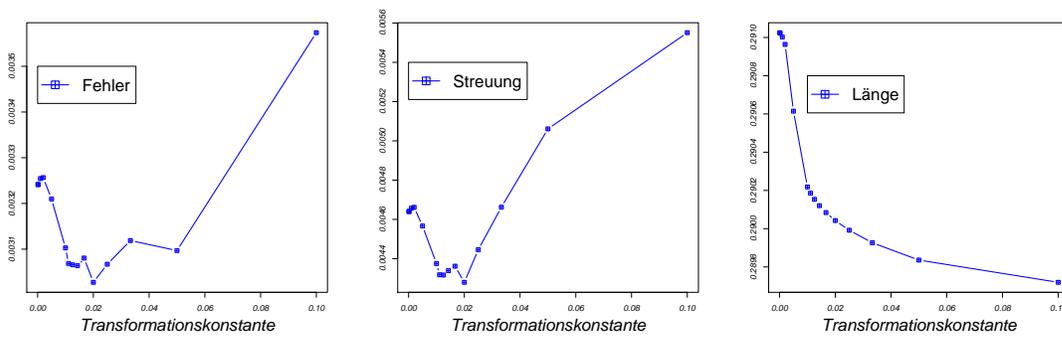


Abbildung 5.2: Ergebnisse der verbundenen k-Optimierung

5.5 Vergleich der Konfidenzintervalle

Verglichen werden nun das neu hergeleitete transformierte Intervall mit bekannten sehr guten Intervallen aus der Literatur. Dazu wird auch das Ausgangsintervall (Wald) mit in die Analyse einbezogen um den Grad der Verbesserung ersichtlich zu gestalten.

5.5.1 Coverage-Kurven

In den Abbildungen 5.3 und 5.4 werden auf $p_2 = 0.5$ bedingte Coverage-Kurven als Funktion von p_1 der untersuchten Intervalle bei einem Stichprobenumfang von $n = 60$ gezeigt. Es wurden zwei verschiedene Odds-Ratios vorgegeben ($\theta = 1, 3$). In diesen beiden Abbildungen erkennt man deutlich, dass das transformierte (und optimierte) Intervall die beste Coverage-Probability aufweist.

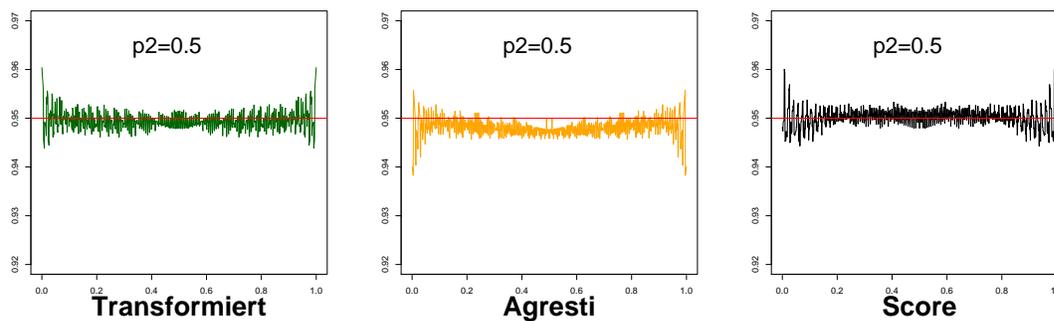


Abbildung 5.3: Coverage-Kurven mit $n=60$ und Odds-Ratio $\theta = 1$ bei gegebenem p_2

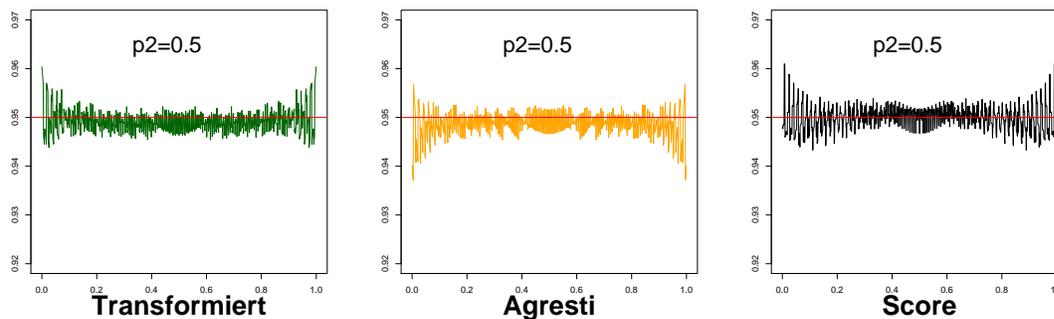


Abbildung 5.4: Coverage-Kurven mit $n=60$ und Odds-Ratio $\theta = 3$ bei gegebenem p_2

Weiterhin zeigt die folgende Tabelle exemplarisch numerische Werte verschiedener Maße.

Tabelle 5.1: Numerische Integrationsergebnisse mit $n=25$ und gegebenem Odds-Ratio $\theta = 3$ und $p_2 = 0.5, 0.1$

p_2	Methode	Coverage	Konservativität	Anti-Konservativität	Fehler	Streuung	Länge
0.5	Wald	0.933	0	0.0172	0.0173	0.0181	0.436
	Agresti	0.947	0.001	0.004	0.005	0.006	0.434
	Transformiert	0.948	0.001	0.003	0.004	0.006	0.440
	Score	0.952	0.004	0.002	0.005	0.007	0.448
0.1	Wald	0.923	0	0.0286	0.028	0.035	0.354
	Agresti	0.954	0.007	0.003	0.01	0.013	0.375
	Transformiert	0.951	0.005	0.005	0.010	0.015	0.368
	Score	0.962	0.012	0.0003	0.013	0.015	0.389

5.6 Numerische Ergebnisse

Tabelle 5.1 zeigt die numerischen Ergebnisse der dargestellten Maße bei einem gegebenem Odds-Ratio $\theta = 3$ und gegebenem $p_2=0.5, 0.1$. Integriert wurde dabei über 10.000 Stützstellen. Aus der Tabelle ist zu erkennen, dass alle Intervalle, außer dem Wald-Intervall, eine sehr gute Coverage-Probability haben.

Die Ergebnisse der numerischen Integrationen werden grafisch in Abbildung 5.6 dargestellt. Die Grafiken zeigen die Maße für die mittlere Coverage-Probability, Länge, Konservativität, Anti-Konservativität, Fehler und Streuung der Intervalle. Gemittelt wurde dabei über vier Stichprobenumfänge. Diese lauten:

$$n_i = 25, 30, 40, 50.$$

Aufgetragen sind dabei diese Mittelwerte für vier verschiedene Odds-Ratios

$$\theta_j = 1, 2, 3, 4.$$

Integriert wurde auf 10000×10000 Stützstellen.

Aus der Abbildung 5.6 geht deutlich hervor, dass das transformierte Intervall in allen Maßen, außer in der Länge, alle untersuchten Intervalle deutlich dominiert. Die Konservativität des Wald-Intervalls ist gleich Null für alle gegebenen Odds-Ratios θ . Ebenso ist die Anti-Konservativität des transformierten Intervalls nahezu Null für alle θ . Außerdem ist die Konservativität des transformierten Intervalls am geringsten, was durch die äußerst geringe Schwankung bzw. Streuung um die Nominalebene $1 - \alpha$ begründet werden kann. Das Score-Intervall von Tango ist sehr konservativ. Wie auch schon in Abbildung 5.4 zu erkennen ist, ist die Streuung des Score-Intervalls hoch. Das überträgt sich auf den mittleren Fehler. Aus Abbildung 5.4 ist ebenfalls zu erkennen, dass das transformierte Intervall am Rand eine sehr gute Coverage-Probability hat, im Gegensatz zu den anderen Verfahren, die am Rand meist eine große Streuung aufzeigen.

Zusammenfassend läßt sich festhalten, dass das transformierte Intervall unter den untersuchten, der Literatur nach besten Intervallen, das beste Intervall ist und bevorzugt angewendet werden sollte.

Abschließend wird eine 3D-Grafik des Coverage-Prozesses und des Längenparaboloiden des transformierten Intervalls gezeigt. Programmiert wurde dazu das Intervall im Programmpaket 'R'. Das

Zusatzpaket 'lattice' erstellt schließlich sogenannte 'Wireframes' von Matrizen, indem für jeden Eintrag der Matrix ein Punkt im 3D-Koordinatensystem erzeugt wird. Der graue Hintergrund ist dabei (leider) fest in das Paket implementiert. Im Paket 'Scatterplot 3D' dient ein mit Gitternetz-versehener Raum als visuelle Grundlage. Leider kann dieses Paket jedoch keine Matrizen lesen. Auf den Aufwand, den Datensatz in die Form zu verändern, dass je ein Parameter aus p_1 festgehalten wird und p_2 stetig läuft, wurde verzichtet.

Die Resultate zeigt Abbildung 5.5.

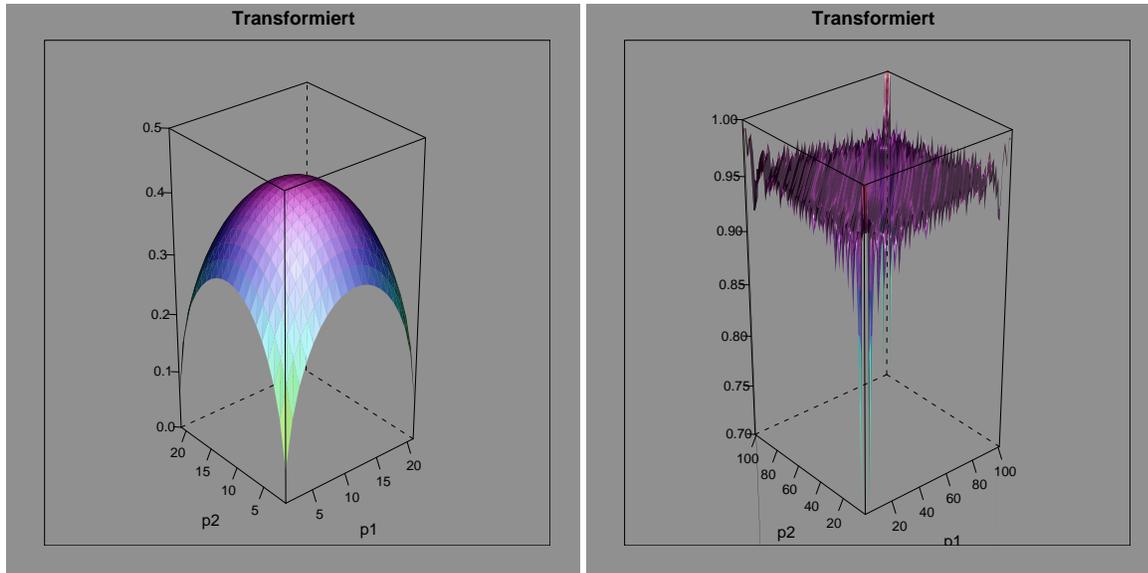


Abbildung 5.5: Längenparaboloid und Coverage-Prozess Transformiert mit $\theta=1$ und $n=40$

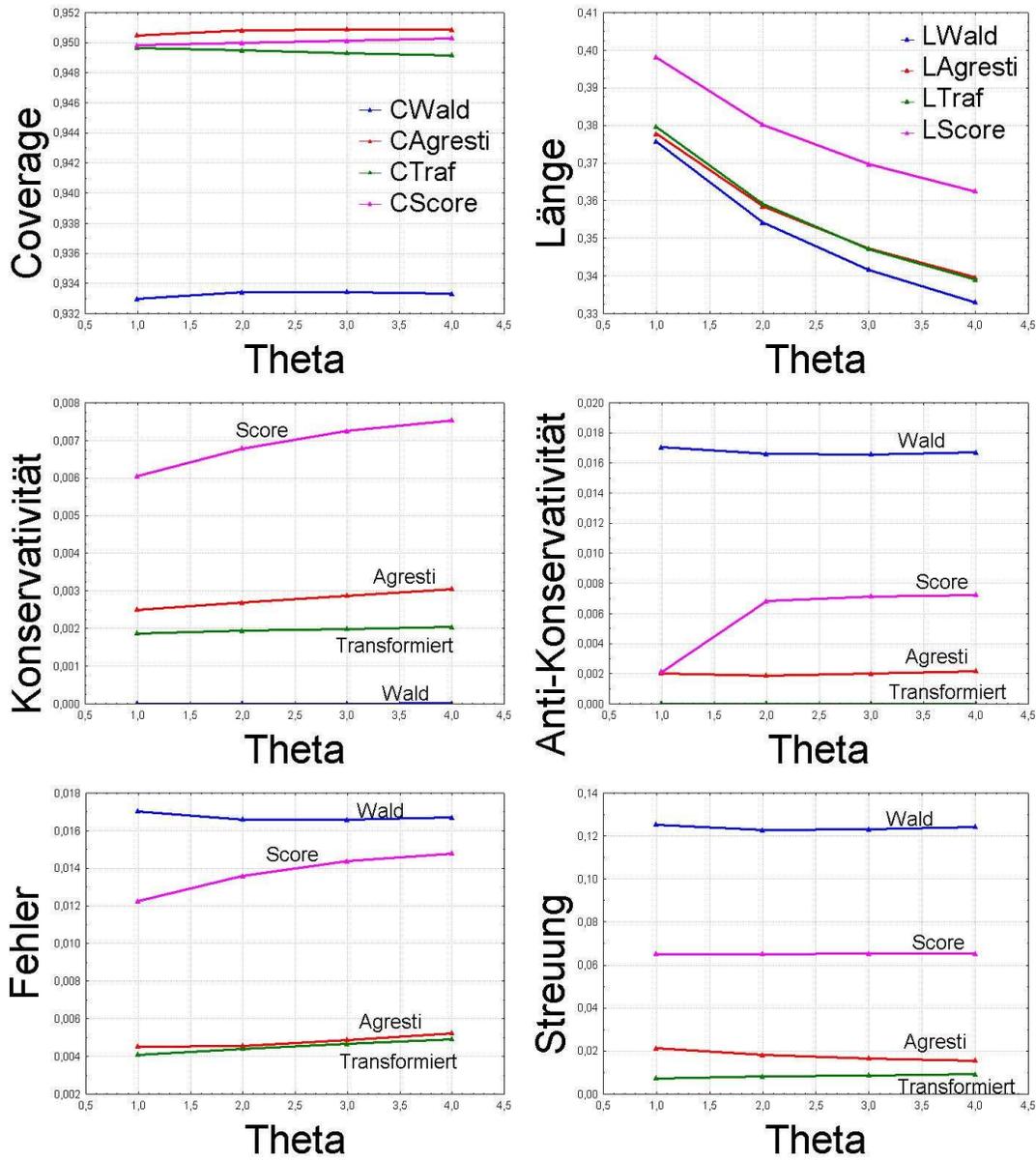


Abbildung 5.6: Vergleich der verbundenen Verfahren

6 Konfidenzintervalle für $p_1 - p_2$ (unverbundene Stichproben)

Untersucht und analysiert werden nun Konfidenzintervalle für die Differenz zweier unverbundener Raten.

6.1 Vereinfachtes Modell mit unabhängigen Stichproben bzw. Methoden

Als Grundlage dient ein Bernoullimodell mit zwei unabhängigen Stichproben. Es seien

$$\begin{aligned} X_{ij} &\stackrel{u.i.v.}{\sim} \mathcal{B}(p_i), \quad i = 1, 2, \quad k = 1, \dots, n_i \\ \Rightarrow Y_1 = \sum_{j=1}^{n_1} X_{1j} &\sim \text{BIN}(n_1, p_1) \quad \text{und} \quad Y_2 = \sum_{j=1}^{n_2} X_{2j} \sim \text{BIN}(n_2, p_2). \end{aligned} \quad (6.1)$$

Die Differenz der Erfolgswahrscheinlichkeiten sei $p_\Delta = p_1 - p_2$.

In Anlehnung an das Einstichprobenproblem sei $\mathcal{KI}(k_1, k_2)$ ein Intervall, welches aus den Erfolgen k_1 und k_2 aus n_1 bzw. n_2 Versuchen berechnet ist. Bevor Statistiken vorgestellt werden, werden zunächst die allgemeinen Maße erläutert. Die Definitionen und teilweise auch die Herleitungen ähneln den verbundenen Stichproben.

6.1.1 Gütemaße

Definition 6.1.1 (UNVERBUNDENE COVERAGE-PROBABILITY)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für $p_\Delta = p_1 - p_2$. Das Maß

$$P(p_\Delta \in \mathcal{KI}(Y_1, Y_2))$$

heißt *unverbundene Coverage-Probability* des Konfidenzintervalls \mathcal{KI} . Die Abbildungen

$$C_{n_1, n_2} : (0, 1)^2 \rightarrow [0, 1], \quad \text{wobei } (p_1, p_2) \mapsto P(p_\Delta \in \mathcal{KI}(Y_1, Y_2)), \quad \text{und} \quad (6.2)$$

$$C_{n_1, n_2, p_2} : (0, 1) \rightarrow [0, 1], \quad \text{wobei } p_1 \mapsto P(p_\Delta \in \mathcal{KI}(Y_1, Y_2)), \quad (6.3)$$

heißen *unverbundener Coverage-Prozess* bzw. *auf p_2 bedingter unverbundener Coverage-Prozess*.

Lemma 6.1.2 (EXAKTE UNVERBUNDENE COVERAGE-PROBABILITY)

Es gelte das Modell (6.1) und es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ . Für die Coverage-Probability gilt:

$$C_{n_1, n_2}(p_1, p_2) = \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} \mathbb{I}_{p_\Delta \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2}. \quad (6.4)$$

Beweis:

$$\begin{aligned}
 P(p_\Delta \in \mathcal{KI}(Y_1, Y_2)) &= C_{n_1, n_2}(p_1, p_2) = E(\mathbb{I}_{p_\Delta \in \mathcal{KI}(Y_1, Y_2)}) \\
 &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} \mathbb{I}_{p_\Delta \in \mathcal{KI}(k_1, k_2)} P(Y_1 = k_1, Y_2 = k_2) \\
 &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} \mathbb{I}_{p_\Delta \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2}.
 \end{aligned}$$

□

Lemma 6.1.3 (EIGENSCHAFTEN DES UNVERBUNDENEN COVERAGE-PROZESSES)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für $p_\Delta = p_1 - p_2$ mit zugehörigem Coverage-Prozess $C_{n_1, n_2}(p_1, p_2)$. Dann gilt:

- (1) $C_{n_1, n_2}(p_1, p_2)$ ist unstetig,
- (2) $C_{n_1, n_2, \frac{1}{2}}(p_1)$ ist symmetrisch um $p_1 = \frac{1}{2}$,
- (3) Es gilt stets $C_{n_1, n_2}(p_1, p_2) = C_{n_2, n_1}(p_2, p_1)$ und damit im Allgemeinen $C_{n_1, n_2}(p_1, p_2) \neq C_{n_2, n_1}(p_1, p_2)$.

Beweis:

- (1) Die Unstetigkeit folgt aus der Unstetigkeit des Indikators.
- (2) Sei $\delta \in (0, \frac{1}{2})$ gegeben, dann gilt:

$$\begin{aligned}
 &C_{n_1, n_2, \frac{1}{2}}\left(\frac{1}{2} + \delta\right) \\
 &= \sum_{k_1}^{n_1} \sum_{k_2}^{n_2} \mathbb{I}_{\delta \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} \left(\frac{1}{2} + \delta\right)^{k_1} \left(\frac{1}{2} - \delta\right)^{n_1-k_1} \left(\frac{1}{2}\right)^{k_2} \left(\frac{1}{2}\right)^{n_2-k_2} \\
 &= \sum_{k_1}^{n_1} \sum_{k_2}^{n_2} \mathbb{I}_{-\delta \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} \left(\frac{1}{2} - \delta\right)^{k_1} \left(\frac{1}{2} + \delta\right)^{n_1-k_1} \left(\frac{1}{2}\right)^{k_2} \left(\frac{1}{2}\right)^{n_2-k_2} \\
 &= C_{n_1, n_2, \frac{1}{2}}\left(\frac{1}{2} - \delta\right).
 \end{aligned}$$

(3)

$$\begin{aligned}
 C_{n_1, n_2}(p_1, p_2) &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} \mathbb{I}_{p_1 - p_2 \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2} \\
 &= \sum_{k_2=0}^{n_2} \sum_{k_1=0}^{n_1} \mathbb{I}_{p_2 - p_1 \in \mathcal{KI}(k_1, k_2)} \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2} \\
 &= C_{n_2, n_1}(p_2, p_1).
 \end{aligned}$$

□

Aus Lemma 6.1.3.3 folgt direkt, dass die zu analysierenden Stichproben stets geordnet werden müssen um aussagefähige Ergebnisse zu bekommen. Orientiert man sich also nicht an einer gegebenen Größenrelation der Stichprobenumfänge, dann sind demnach die erzeugten Coverage-Kurven falsch. Um gute Ergebnisse zu bekommen, sei deshalb vereinbart:

Die Stichprobenumfänge erfüllen stets folgende Relation:

$$n_1 \geq n_2.$$

Es werden nur Konfidenzintervalle für

$$p_1 - p_2 = p_\Delta$$

betrachtet und analysiert.

Abbildung 6.1: Anforderungen an die Stichprobenumfänge bei unverbundenen Stichproben

Die Symmetrieeigenschaft überträgt sich natürlich auf den unbedingten Prozess. Damit wird auch hier der Programmieraufwand erleichtert.

Definition 6.1.4 (MITTLERE UNVERBUNDENE COVERAGE-PROBABILITY)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehörigem Coverage-Prozess $C_{n_1, n_2}(p_1, p_2)$. Die Größen

$$\overline{CU}_{n_1, n_2} = \int_0^1 \int_0^1 C_{n_1, n_2}(p_1, p_2) dp_1 dp_2 \quad \text{und} \quad (6.5)$$

$$\overline{CU}_{n_1, n_2, p_2} = \int_0^1 C_{n_1, n_2, p_2}(p_1) dp_1 \quad (6.6)$$

heißen mittlere unverbundene Coverage-Probability und mittlere unverbundene Coverage-Probability für gegebenes p_2 des Konfidenzintervalls \mathcal{KI} .

Definition 6.1.5 (UNVERBUNDENE LÄNGENKURVE)

Es gelte das Modell (6.1) und es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ . Die Abbildungen

$$L_{n_1, n_2} : (0, 1)^2 \rightarrow [0, 1], \text{ wobei } (p_1, p_2) \mapsto E(O(Y_1, Y_2) - \mathcal{U}(Y_1, Y_2)), \text{ und} \quad (6.7)$$

$$L_{n_1, n_2, p_2} : (0, 1) \rightarrow [0, 1], \text{ wobei } p_1 \mapsto E_{p_2}(O(Y_1, Y_2) - \mathcal{U}(Y_1, Y_2)), \quad (6.8)$$

heißen unverbundene Längenkurve und unverbundene Längenkurve mit gegebenem p_2 des Konfidenzintervalls \mathcal{KI} .

Anmerkung 6.1.6 Analog bedeutet die Notation E_{p_2} , dass der auf p_2 bedingte Erwartungswert betrachtet wird.

Lemma 6.1.7 (ERWARTETE UNVERBUNDENE LÄNGE)

Es gelte das Modell (6.1) und es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehöriger Längenkurve $L_{n_1, n_2}(p_1, p_2)$. Für die erwartete Länge gilt:

$$L_{n_1, n_2}(p_1, p_2) = \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2}.$$

Beweis:

Es gilt:

$$\begin{aligned} & E [O(Y_1, Y_2) - \mathcal{U}(Y_1, Y_2)] \\ &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2}. \end{aligned}$$

□

Definition 6.1.8 (MITTLERE UNVERBUNDENE LÄNGE)

Es gelte das Modell (6.1). Es sei $\mathcal{K}I(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehöriger Längenkurve $L_{n_1, n_2}(p_1, p_2)$. Die Größen

$$\overline{LU}_{n_1, n_2} = \int_0^1 \int_0^1 L_{n_1, n_2}(p_1, p_2) dp_1 dp_2 \quad \text{und} \quad (6.9)$$

$$\overline{LU}_{n_1, n_2, p_2} = \int_0^1 L_{n_1, n_2}(p_1) dp_1 \quad (6.10)$$

heißen mittlere unverbundene Länge und mittlere unverbundene Länge für gegebenes p_2 des Konfidenzintervalls $\mathcal{K}I$.

Lemma 6.1.9 (EIGENSCHAFTEN DER UNVERBUNDENEN LÄNGENKURVE)

Es gelte das Modell (6.1). Es sei $\mathcal{K}I(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehöriger Längenkurve $L_{n_1, n_2}(p_1, p_2)$. Dann gilt:

- (1) $L_{n_1, n_2}(p_1, p_2)$ ist auf $(0, 1)^2$ total differenzierbar und damit stetig.
- (2) $L_{n_1, n_2, \frac{1}{2}}(p_1)$ ist symmetrisch um $p_1 = \frac{1}{2}$.
- (3) $L_{n_1, n_2}(p_1, p_2)$ beschreibt einen nach unten geöffneten Paraboloiden mit Maximum in $L_{n_1, n_2}\left(\frac{1}{2}, \frac{1}{2}\right)$.

Beweis:

- (1) Die partielle Differenzierbarkeit folgt aus der Existenz der partiellen Ableitungen der einzelnen Summanden. Schließlich gilt:

$$\begin{aligned} \frac{\partial L_{n_1, n_2}}{\partial p_1} &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \binom{n_1}{k_1} \binom{n_2}{k_2} p_2^{k_2} (1-p_2)^{n_2-k_2} \times \\ &\quad \left[p_1^{k_1-1} (1-p_1)^{n_1-k_1-1} (k_1(1-p_1) - p_1(n_1-k_1)) \right], \\ \frac{\partial L_{n_1, n_2}}{\partial p_2} &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \binom{n_1}{k_1} \binom{n_2}{k_2} p_1^{k_1} (1-p_1)^{n_1-k_1} \times \\ &\quad \left[p_2^{k_2-1} (1-p_2)^{n_2-k_2-1} (k_2(1-p_2) - p_2(n_2-k_2)) \right]. \end{aligned}$$

Aus der Stetigkeit der partiellen Ableitungen folgt die totale Differenzierbarkeit.

(2) Es seien $\delta \in (0, \frac{1}{2})$ und p_2 gegeben, dann gilt:

$$\begin{aligned}
 L_{n_1, n_2, \frac{1}{2}} \left(\frac{1}{2} + \delta \right) &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \times \\
 &\quad \binom{n_1}{k_1} \binom{n_2}{k_2} \left(\frac{1}{2} + \delta \right)^{k_1} \left(1 - \left(\frac{1}{2} + \delta \right) \right)^{n_1 - k_1} \left(\frac{1}{2} \right)^{k_2} \left(\frac{1}{2} \right)^{n_2 - k_2} \\
 &= \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} [O(k_1, k_2) - \mathcal{U}(k_1, k_2)] \times \\
 &\quad \binom{n_1}{k_1} \binom{n_2}{k_2} \left(\frac{1}{2} - \delta \right)^{k_1} \left(\frac{1}{2} + \delta \right)^{n_1 - k_1} \left(\frac{1}{2} \right)^{k_2} \left(\frac{1}{2} \right)^{n_2 - k_2} \\
 &= L_{n_1, n_2, \frac{1}{2}} \left(\frac{1}{2} - \delta \right).
 \end{aligned}$$

(3) Der Beweis verläuft analog zum Beweis des Maximums der Längenkurve im Einstichprobenproblem. \square

Definition 6.1.10 (MITTLERE UNVERBUNDENE KONSERVATIVITÄT und ANTI-KONSERVATIVITÄT)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehörigem unverbundenen Coverage-Prozess $C_{n_1, n_2}(p_1, p_2)$. Die Größen

$$\overline{KU}_{n_1, n_2} = \int \int_{\{C_{n_1, n_2}(p_1, p_2) \geq 1 - \alpha\}} C_{n_1, n_2}(p_1, p_2) dp_1 dp_2 \quad \text{und} \quad (6.11)$$

$$\overline{KU}_{n_1, n_2, p_2} = \int_{\{C_{n_1, n_2, p_2}(p_1) \geq 1 - \alpha\}} C_{n_1, n_2, p_2}(p_1) dp_1 \quad \text{bzw.} \quad (6.12)$$

$$\overline{AKU}_{n_1, n_2} = \int \int_{\{C_{n_1, n_2}(p_1, p_2) \leq 1 - \alpha\}} C_{n_1, n_2}(p_1, p_2) dp_1 dp_2 \quad \text{und} \quad (6.13)$$

$$\overline{AKU}_{n_1, n_2, p_2} = \int_{\{C_{n_1, n_2, p_2}(p_1) \leq 1 - \alpha\}} C_{n_1, n_2, p_2}(p_1) dp_1 \quad (6.14)$$

heißen unverbundene mittlere Konservativität und unverbundene mittlere Konservativität für gegebenes p_2 bzw. unverbundene mittlere Anti-Konservativität und unverbundene mittlere Anti-Konservativität für gegebenes p_2 des Intervalls \mathcal{KI} .

Analog zum Einstichprobenproblem werden diese Maße mit Hilfe der Coverage-Kurven des Intervalls \mathcal{KI} berechnet und als die Gesamtfläche (Gesamtvolumen) zwischen der Coverage-Kurve und des Konfidenzniveaus angesehen.

Definition 6.1.11 (MITTLERER UNVERBUNDENER FEHLER BZW. DISTANZ)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehörigem Coverage-Prozess $C_{n_1, n_2}(p_1, p_2)$. Die Größen

$$\overline{FU}_{n_1, n_2} = \int_0^1 \int_0^1 |C_{n_1, n_2}(p_1, p_2) - (1 - \alpha)| dp_1 dp_2 \quad \text{und} \quad (6.15)$$

$$\overline{FU}_{n_1, n_2, p_2} = \int_0^1 |C_{n_1, n_2, p_2}(p_1) - (1 - \alpha)| dp_1 \quad (6.16)$$

heißen unverbundener mittlerer Fehler des Intervalls \mathcal{KI} und unverbundener mittlerer Fehler für gegebenes p_2 des Konfidenzintervalls \mathcal{KI} .

Definition 6.1.12 (MITTLERE STREUUNG)

Es gelte das Modell (6.1). Es sei $\mathcal{KI}(Y_1, Y_2)$ ein Konfidenzintervall für p_Δ mit zugehörigem Coverage-Prozess $C_{n_1, n_2}(p_1, p_2)$. Die Größen

$$\overline{SU}_{n_1, n_2} = \sqrt{\int_0^1 \int_0^1 (C_{n_1, n_2}(p_1, p_2) - (1 - \alpha))^2 dp_1 dp_2} \quad \text{und} \quad (6.17)$$

$$\overline{SU}_{n_1, n_2, p_2} = \sqrt{\int_0^1 (C_{n_1, n_2, p_2}(p_1) - (1 - \alpha))^2 dp_1} \quad (6.18)$$

heißen unverbundene mittlere Streuung des Prozesses des Intervalls bzw. unverbundene mittlere Streuung mit gegebenem p_2 des Intervalls \mathcal{KI} .

6.2 Konfidenzintervalle für $p_\Delta = p_1 - p_2$

Die Intention zur Herleitung und Beschreibung der Intervalle lehnt sich eng an das verbundene Zweistichprobenproblem an. Es sind daher nur die notwendigsten Ausgangspunkte aufgelistet. Ausgangspunkt sei das Modell zur Beschreibung von zwei unabhängigen dichotomen Stichproben. Ein asymptotisches Konfidenzintervall für $p_1 - p_2$ ist gegeben durch

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (6.19)$$

und wird analog zum Einstichprobenproblem als das Wald-Konfidenzintervall bezeichnet (vgl. [19]). Die Coverage-Probability dieses Konfidenzintervalls ist äußerst gering, besonders für kleine Stichprobenumfänge. In der Arbeit von AGRESTI und CAFFO (2000) [19] wurde ein Verfahren veröffentlicht, welches auf dem heuristischen Ansatz der Konstruktion des Agresti Intervalls für einen Parameter aufbaut. Die Idee dazu soll sein, auf den Gesamtstichprobenumfang N eine gewisse Anzahl von Erfolgen und Misserfolgen zu addieren und damit die Schätzer der Erfolgswahrscheinlichkeiten zu verändern. Das neue Intervall ergibt sich schließlich, indem diese Schätzer in die Gleichung des Wald-Intervalls eingesetzt werden. Das sogenannte Agresti-Konfidenzintervall (Wald+t-Intervall) hat damit folgende Gestalt:

$$\tilde{p}_1 - \tilde{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}, \quad (6.20)$$

$$\text{wobei} \quad \tilde{p}_i = \frac{k_i + \frac{t}{4}}{n_i + \frac{t}{2}}, \quad \tilde{n}_i = n_i + \frac{t}{2}, \quad i = 1, 2,$$

gilt. Dieses Intervall ist, wie auch das Wald-Intervall, nicht bereichserhaltend. Ragen die Ober- bzw. Untergrenze über +1 bzw. -1 hinaus, dann werden den Grenzen der Wert +1 bzw. -1 zugewiesen. AGRESTI und CAFFO (2000) [19] kamen zu dem Ergebnis, dass insgesamt vier Beobachtungen auf den Stichprobenumfang addiert werden- zwei Erfolge und zwei Mißerfolge. Die Herleitung folgt auch hier einem heuristischen Verfahren, indem die Wald+t Intervalle für verschiedene Maße verglichen werden.

NEWCOMBE (1998) [30] untersuchte elf verschiedene Verfahren zur Konstruktion von Konfidenzintervallen für die Differenz zweier unabhängiger Erfolgswahrscheinlichkeiten. In der Arbeit wurde ein Verfahren vorgestellt, dass das Wald-Konfidenzintervall erheblich verbessert und ähnlich zum Wilson-Intervall konstruiert wird. Auch dieses neue Intervall entsteht aus der Invertierung

eines Score-Tests. Genauer gesagt ist das sogenannte 'Hybrid-Score'-Intervall von Newcombe ein Hybrid (so wie der Name bereits verrät) bzw. eine Mischung aus den Ergebnissen des Einstichprobenproblems.

Es seien

$$(l_i, u_i) \quad (6.21)$$

die Lösungen in p_i für

$$\frac{|\hat{p}_i - p_i|}{\sqrt{p_i(1-p_i)/n_i}} = z_{\alpha/2}. \quad (6.22)$$

Newcombes Hybrid-Score Intervall lautet schließlich:

$$\left[\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} \sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{u_2(1-u_2)}{n_2}}; \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} \sqrt{\frac{u_1(1-u_1)}{n_1} + \frac{l_2(1-l_2)}{n_2}} \right]. \quad (6.23)$$

Dieses Intervall hat die Eigenschaften, dass es bei auftretenden Null- oder Vollergebnissen an den Rändern nicht degeneriert. Damit ist dieses Intervall auch für hohe- bzw. kleine Effekte einsetzbar.

6.3 Transformierte Statistiken

Die transformierten Intervalle lehnen sich eng an das verbundene Zweistichprobenproblem an. Auch bei unverbundenen Linearkombinationen der Wahrscheinlichkeiten werden diese mit der Fisher-Funktion transformiert. Analog müssen die Null- und Vollergebnisse abgefangen werden. Dieses geschieht durch den Einbau einer Konstanten k in die Transformationsfunktion. Die Eigenschaften und Folgerungen aus der Konstanten übertragen sich entsprechend aus Kapitel 5.4.2. Ausserdem wird auch hier der Varianzschätzer durch eine Nullfolge modifiziert, da die transformierten Intervalle sonst leicht anti-konservativ sind. Das monotone Verhalten der mittleren Länge der Konfidenzschär überträgt sich aus dem verbundenen Zweistichprobenproblem. Die Nullfolge muss im unverbundenen Fall aus zwei Stichprobenumfängen konstruiert sein. Um schließlich ein analoges Vorgehen zum verbundenen Zweistichprobenproblem aufzeigen zu können, wird bei dieser $n = n_1 + n_2$ festgelegt und diese Beziehung in die Konstruktion der Nullfolge aus (5.23) eingesetzt.

Lemma 6.3.1 (TRANSFORMIERTE INTERVALLE)

Es gelte das Modell (6.1). Es sei \hat{p}_i , $i=1,2$, die relative Häufigkeit der jeweiligen Stichprobe und es sei $\hat{p}_\Delta = \hat{p}_1 - \hat{p}_2$ die Linearkombination dieser erwartungstreuen Schätzer. Weiterhin sei $c(n_1, n_2) = \frac{v}{(n_1+n_2)^k}$ eine Nullfolge gemäß (5.23). Durch die Betrachtung und die Anwendung der Funktionenscharen

$$g_k : (-1, 1) \rightarrow \mathbb{R}, \quad (p_1 - p_2) = p_\Delta \mapsto \frac{1}{2} \cdot \log \left(\frac{1+k+p_\Delta}{1+k-p_\Delta} \right), \quad (6.24)$$

mit den Ableitungen

$$\frac{dg_k(p_\Delta)}{dp_\Delta} = \frac{1+k}{(1+k)^2 - p_\Delta^2} \quad (6.25)$$

und den Umkehrfunktionen

$$g^{-1} : \mathbb{R} \rightarrow (0, 1), \quad x \mapsto \frac{\exp(2x) - 1}{\exp(2x) + 1}, \quad (6.26)$$

dann sind durch

$$P(g^{-1}(\mathcal{U}^T) \leq p_1 - p_2 \leq g^{-1}(\mathcal{O}^T)) \approx 1 - \alpha, \quad (6.27)$$

wobei

$$\mathcal{U}^T, \mathcal{O}^T = g_k(\hat{p}_\Delta) \pm z_{1-\alpha/2} \cdot \frac{dg_k(\hat{p}_\Delta)}{d\hat{p}_\Delta} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + c(n_1, n_2)}$$

gilt, asymptotische und bereichserhaltende $(1 - \alpha)$ -Konfidenzintervalle für p_Δ gegeben.

Beweis:

Der Beweis verläuft analog zu der Herleitung des transformierten Konfidenzintervall für die verbundene Differenz $p_1 - p_2$ (siehe Lemma 5.4.5). \square

Wie bereits erwähnt, bewirken die Parametrisierungsparameter eine Veränderung der Coverage-Probability. Es muss ein spezielles Intervall aus der Schar ausgewählt werden. Diese Auswahl muss analytisch begründet werden. Die Auswahl eines Intervalls wird, analog zum verbundenen Design, durch eine multikriterielle Optimierung hergeleitet. Die Existenz der Minima sichert die Monotonie der mittleren Längenkurven (vgl. Satz 5.4.6).

Satz 6.3.2 (OPTIMALE PARAMETERKOMBINATION)

Es gelte das Modell (6.1). Es sei $\mathcal{K}I_{k,v,\lambda}(Y_1, Y_2)$ das transformierte und parametrisierte Konfidenzintervall in Abhängigkeit der Transformationskonstante k mit zugehörigem Coverage-Prozess $C_{n_1, n_2, k, v, \lambda}$ und eingebauter Nullfolge $c(n_1, n_2) = \frac{v}{(n_1 + n_2)^\lambda}$. Die Größen $\overline{SU}_{n_1, n_2, k, v, \lambda}$ bzw. $\overline{FU}_{n_1, n_2, k, v, \lambda}$ seien die mittlere Streuung bzw. der mittlere Fehler in Abhängigkeit der Parameter aus den Definitionen¹ 6.1.12 bzw. 6.1.11. Im Mittel führt die Kombination

$$(k^*, v^*, \lambda^*) = \left(\frac{1}{50}, 1, 2 \right) \quad (6.28)$$

zu einer lokalen minimalen Schwankung des zugehörigen Coverage-Prozesses $C_{n_1, n_2, k, v, \lambda}(p_1, p_2)$ um die $(1 - \alpha)$ -Ebene.

Analog zum verbundenen Design wird bei der Optimierung ein Notationswechsel vorgenommen. In der Optimierung geht es darum, dass Verhalten der Coverage-Probability in Abhängigkeit der Parametrisierungsparameter zu analysieren. Es wird daher der mittlere Fehler bzw. die mittlere Streuung der Konfidenzschar als Funktion der Bedingungsparameter angesehen. Es gilt also

$$\begin{aligned} \overline{SU}_{n_1, n_2, k, v, \lambda} &= \overline{SU}(n_1, n_2, k, v, \lambda), \\ \overline{FU}_{n_1, n_2, k, v, \lambda} &= \overline{FU}(n_1, n_2, k, v, \lambda). \end{aligned}$$

Herleitung:

Das Vorgehen zur Bestimmung der optimalen Intervalle erfolgt analog zum Bilevel-Algorithmus

¹Die Aufnahme der Größen k, v, λ in den Index bedeutet, dass auf diesen numerischen Größen bedingt wird. Diese werden also fest vorgegeben.

aus dem verbundenen Zweistichprobenproblem. Natürlich muss der Algorithmus verschiedene Stichprobenumfänge enthalten. Zugelassen sind sowohl gleiche, als auch ungleiche Stichprobenumfänge. Vorsicht sei allerdings bei der Wahl der Stichprobenumfänge geboten. Der Optimierungsalgorithmus muss so programmiert sein, dass die Bedingung aus Abbildung 6.1 nicht verletzt wird.

Analog zum Aufbau des Programms aus Satz 5.4.8 gilt es nun folgenden Optimierungsalgorithmus zu lösen:

$$Z_1 = \frac{1}{N_1 N_2} \underbrace{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{SU}(n_{1i}, n_{2j}, k^*, v, \lambda)}_{n_{1j} \geq n_{2j}} \rightarrow \min_{v, \lambda}, \quad (6.29)$$

unter den Nebenbedingungen:

$$k^* = \min_k \frac{1}{N_1 N_2} \underbrace{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{SU}(n_{1i}, n_{2j}, k, v, \lambda)}_{n_{1j} \geq n_{2j}}, \quad (6.30)$$

$v, \lambda, k > 0$,

$$Z_2 = \frac{1}{N_1 N_2} \underbrace{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{FU}(n_{1i}, n_{2j}, k^*, v, \lambda)}_{n_{1j} \geq n_{2j}} \rightarrow \min_{v, \lambda}, \quad (6.31)$$

unter den Nebenbedingungen:

$$k^* = \min_k \frac{1}{N_1 N_2} \underbrace{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{FU}(n_{1i}, n_{2j}, k, v, \lambda)}_{n_{1j} \geq n_{2j}}, \quad (6.32)$$

$v, \lambda, k > 0$.

Ergebnis:

Als Stützstellen wurden diese Stichprobenumfänge übergeben:

- Stichprobenumfänge n_{1i} der ersten Stichprobe, wobei $i = 1, \dots, N_1$ gilt:

$$n_{1i} = 35, 40, 45, 50.$$

- Stichprobenumfänge n_{2j} der zweiten Stichprobe, wobei $j = 1, \dots, N_2$ gilt:

$$n_{2j} = 20, 25, 30, 35.$$

Mit diesen Stützstellen wurde das folgende Resultat erzielt:

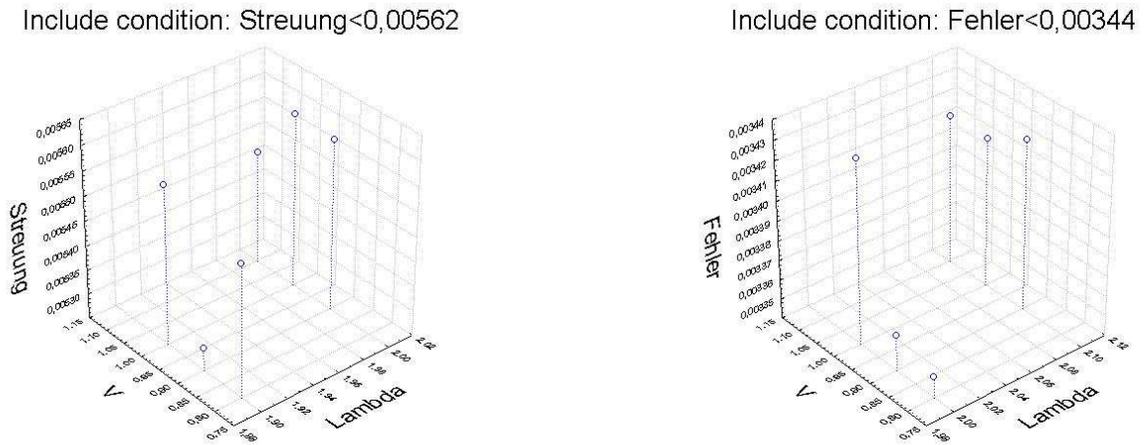


Abbildung 6.2: Ergebnismenge der unverbundenen Optimierung

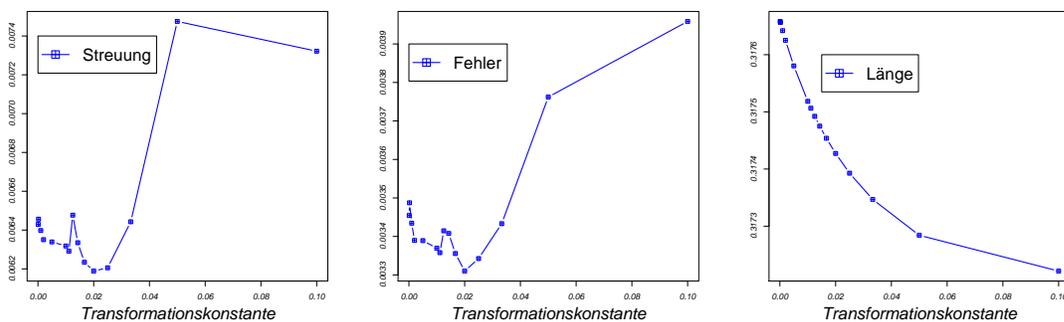


Abbildung 6.3: Ergebnisse der unverbundenen k-Optimierung

6.4 Vergleich der Konfidenzintervalle

Der Vergleich der vorgestellten Verfahren erfolgt sowohl grafisch, als auch anhand mathematischer Maße. Zunächst werden die Coverage-Prozesse der Konfidenzintervalle in den Abbildungen 6.4 und 6.5 dargestellt. Bedingt wurde hier auf $p_2 = 0.1, 0.5$. Ausserdem wurden dabei gleiche Stichprobenumfänge mit $n_1 = n_2 = 40$ verwendet.

6.4.1 Coverage-Prozesse

Es werden zunächst die unverbundenen Coverage-Prozesse der untersuchten Konfidenzintervalle in den Abbildungen 6.4 und 6.5 gezeigt. Aus Platzgründen bzw. aus Gründen der Übersichtlichkeit wird das Wald-Konfidenzintervall für unverbundene Stichproben im Vergleich der Coverage-Kurven nicht gezeigt.

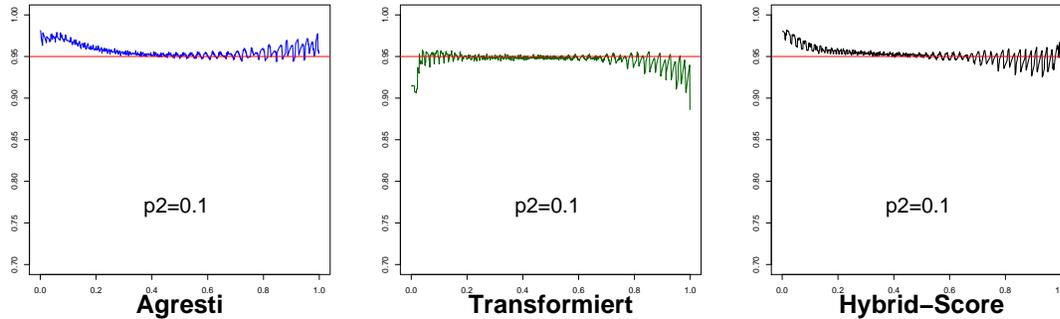


Abbildung 6.4: Coverage-Kurven bedingt auf $p_2=0.1$ mit $n_1 = n_2 = 40$

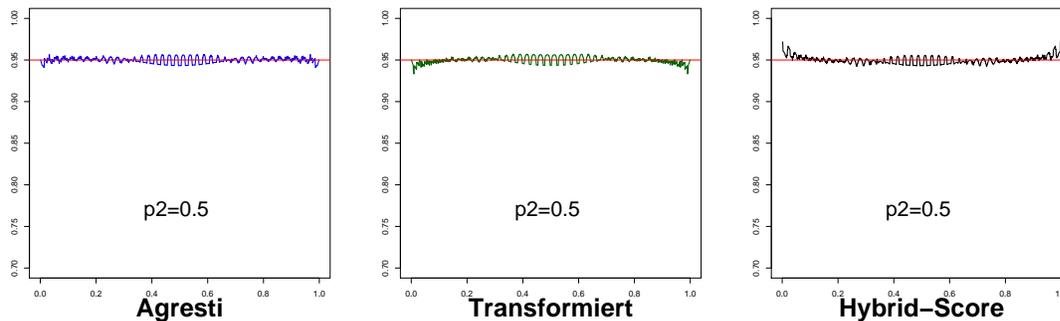


Abbildung 6.5: Coverage-Kurven bedingt auf $p_2=0.5$ mit $n_1 = n_2 = 40$

Weiterhin wurden die vorgestellten Maße untersucht. Die Ergebnisse zeigt Abbildung 6.6. Die Abbildungen 6.4 und 6.5 zeigen deutlich das konservative Verhalten des Agresti- und des Hybrid-Score Intervalls auf. Besonders für $|p_1 - p_2| \approx 0$ sind diese Intervalle äußerst konservativ. Das Wald-, sowie das transformierte Intervall sind in diesem Bereich hingegen eher liberal. Weiterhin werden für verschiedene Stichprobenumfänge die Intervalle bewertet und anhand der vorgestellten mathematischen Maße verglichen. Die Ergebnisse dazu zeigt die Abbildung 6.6. In dieser wurde der gemittelte Stichprobenumfang aus n_1 und n_2 auf der Abszisse gegen das bivariate numerische Integrationsergebnis aufgetragen.

Als Gesamtergebnis läßt sich festhalten, dass keines der Intervalle als das beste ausgewiesen werden kann, denn die Ergebnisse überkreuzen sich. Die Unterschiede zwischen den Ergebnissen liegen in der Größenordnung von 1×10^{-4} . Vielleicht sollte festgehalten werden, dass das transformierte Intervall erst ab einem mittleren Stichprobenumfang mit $n \equiv 25$ angewendet werden sollte.

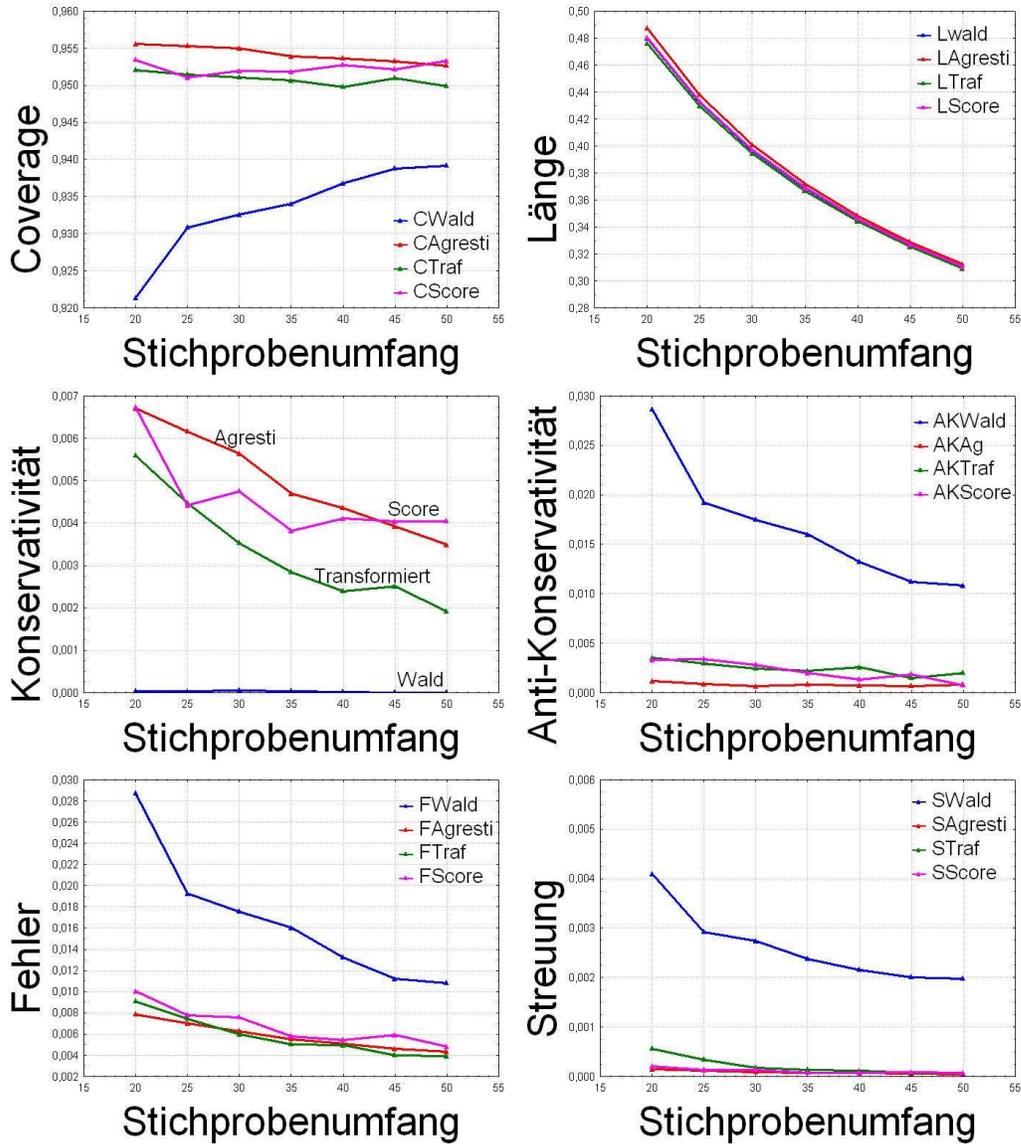


Abbildung 6.6: Vergleich der unverbundenen Verfahren

7 Auswertung des Beispiels

Ausgewertet wird nun das kleine Beispiel zu einer Diagnosestudie mit verbundenen Stichproben. Es werden die Grenzen der Konfidenzintervalle, sowie p-Werte zu den Hypothesentests $H_0 : \pi_1 = \pi_2$, angegeben.

- Konfidenzintervall für $\pi_2 - \pi_1$:

Intervall	Untergrenze	Obergrenze
Wald	0.10	0.22
Agresti	0.10	0.22
Transformiert	0.11	0.22

- p-Werte für $H_0 : \pi_1 = \pi_2$:

Test	p-Wert
Wald	<0.0001
Agresti	<0.0001
Transformiert	<0.0001

Die Grenzen decken sich hier bei allen drei dargestellten Verfahren, was allerdings durch den sehr hohen Stichprobenumfang begründet werden kann. Weiterhin besteht ein signifikanter Unterschied zwischen der Diagnose des erfahrenen Radiologen und dem computerunterstützten Verfahren (CAD). Es handelt sich hier um einen realen Datensatz. Aus den Ergebnissen ist unter anderem zu folgern, dass sich die moderne Medizin im Bereich der Diagnostik noch sehr stark verbessern muss.

8 Zusammenfassung und Ausblick

In dieser Arbeit wurde eine wissenschaftliche Analyse von Konfidenzintervallen für Linearkombinationen von Erfolgswahrscheinlichkeiten vorgenommen. Als Grundlage dieser Arbeit diente das sogenannte Wald-Konfidenzintervall für eine Erfolgsrate, dessen schlechte Eigenschaften in der Arbeit gezeigt worden sind. Das Ziel der Arbeit sollte sein, ein Verfahren zu konstruieren, mit welchem sich einfache und gute Konfidenzintervalle für Linearkombinationen von Erfolgswahrscheinlichkeiten berechnen lassen.

Das hier vorgestellte Verfahren orientiert sich an der Transformationsmethode und wurde auf Linearkombinationen von Erfolgswahrscheinlichkeiten erweitert. Die hergeleiteten Verfahren stellen Scharen von Konfidenzintervallen dar, die schließlich multikriteriell optimiert wurden.

Die Bewertung der Konfidenzintervalle erfolgte rein mathematisch mit bekannten und neu konstruierten Maßen. Diese Maße wurden genauestens vorgestellt und anhand von Berechnungsbeispielen und Grafiken erklärt. Die Vergleiche der Verfahren, sowie die Optimierungsalgorithmen, erfolgten mit Hilfe bivariater numerischer Integrationen. Angewendet wurde die zusammengesetzte Trapez-Regel. Die auf diese Weise hergeleiteten Statistiken zeigten sowohl im Einstich- sowie im Zweistichprobenproblem gute Ergebnisse.

Zusammenfassend wurde in der Arbeit eine Möglichkeit eingeführt, mit welcher sich sehr einfach Konfidenzintervalle für Linearkombinationen von Erfolgswahrscheinlichkeiten konstruieren lassen. Außerdem ist diese Methode in elementare Vorlesungen über statistische Datenanalyse integrierbar. Vielleicht ist die Unterrichtung dieser Intervalle leichter als die der Agresti-Intervalle, aufgrund der mathematischen Herleitung.

Die Transformationsmethode ist weiterhin auf mehrfaktorielle Anlagen mit dichotomen Daten übertragbar, die häufig in Diagnosestudien auftreten. Sei also ein Modell einer Diagnosestudie mit mehreren Methoden und einem Reader mit Mehrfachmessung gegeben. Es sei sichergestellt, dass kein Patient sowohl gesund als auch krank ist. Die Anzahl der Methoden sei d . Dann sind die bekannten Wald-Typ- oder ANOVA-Typ-Statistiken aufstellbar.

Betrachtet man schließlich die mehrdimensionalen Funktionen

$$\begin{aligned} L: (0, 1)^d &\rightarrow \mathbb{R}^d, & (w_1, w_2, \dots, w_d)^t &\mapsto (\text{Logit}(w_1), \text{Logit}(w_2), \dots, \text{Logit}(w_d))^t, \\ P: (0, 1)^d &\rightarrow \mathbb{R}^d, & (w_1, w_2, \dots, w_d)^t &\mapsto (\text{Probit}(w_1), \text{Probit}(w_2), \dots, \text{Probit}(w_d))^t, \end{aligned}$$

dann erhält man folgende Jacobi-Matrizen:

$$\begin{aligned} [\mathbf{L}(\mathbf{w})]' &= \frac{\partial \mathbf{L}}{\partial \mathbf{w}} = \begin{pmatrix} \frac{1}{w_1(1-w_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{w_2(1-w_2)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{w_d(1-w_d)} \end{pmatrix} \in \mathbb{R}^{d \times d}, \\ [\mathbf{P}(\mathbf{w})]' &= \frac{\partial \mathbf{P}}{\partial \mathbf{w}} = \begin{pmatrix} \frac{\sqrt{2\pi}}{\exp\{-1/2w_1^2\}} & 0 & \cdots & 0 \\ 0 & \frac{\sqrt{2\pi}}{\exp\{-1/2w_2^2\}} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sqrt{2\pi}}{\exp\{-1/2w_d^2\}} \end{pmatrix} \in \mathbb{R}^{d \times d}. \end{aligned}$$

Diese Funktionen können als mögliche Transformationsfunktionen in die quadratischen Formen der beiden folgenden Sätze eingesetzt werden.

Satz 8.0.1 (TRANSFORMIERTE WALD-TYP-STATISTIK)

Die Funktion \mathbf{g} genüge den Anforderungen an eine Transformationsfunktion. Es sei außerdem

$$\mathbf{g}(\mathbf{w}) = (g_1(w_1), g_2(w_2), \dots, g_d(w_d))^t \quad \text{bzw.} \quad \mathbf{g}(\hat{\mathbf{w}}) = (g_1(\hat{w}_1), g_2(\hat{w}_2), \dots, g_d(\hat{w}_d))^t$$

der Vektor mit den transformierten (geschätzten) Effekten. Dann gilt:

$$Q_N^*(\mathbf{C}) = N[\mathbf{g}(\hat{\mathbf{w}})]^t \mathbf{C}^t \left(\mathbf{C}[\mathbf{g}'(\hat{\mathbf{w}})] \hat{\mathbf{V}}_N [\mathbf{g}'(\hat{\mathbf{w}})]^t \mathbf{C}^t \right)^{-} \mathbf{C} \mathbf{g}(\hat{\mathbf{w}})$$

ist asymptotisch χ_f^2 verteilt mit $f = \text{rang}(C)$ Freiheitsgraden, wobei $\hat{\mathbf{V}}_N$ einen konsistenten Schätzer der Kovarianzmatrix \mathbf{V}_N bezeichnet.

Satz 8.0.2 (TRANSFORMIERTE ANOVA-TYP-STATISTIK)

Es sei \mathbf{C} eine Kontrastmatrix und $\mathbf{T} = \mathbf{C}^t (\mathbf{C} \mathbf{C}^t)^{-} \mathbf{C}$, wobei $(\mathbf{C} \mathbf{C}^t)^{-}$ eine beliebige verallgemeinerte Inverse zu $(\mathbf{C} \mathbf{C}^t)$ ist. Es gelte außerdem $\text{Spur}(\mathbf{T}[\mathbf{g}'(\mathbf{w})] \hat{\mathbf{V}}_N [\mathbf{g}'(\mathbf{w})]^t) \neq 0$. Dann kann

$$F_N(\mathbf{C}) = \frac{N}{\text{Spur}(\mathbf{T}[\mathbf{g}'(\hat{\mathbf{w}})] \hat{\mathbf{V}}_N [\mathbf{g}'(\hat{\mathbf{w}})]^t)} [\mathbf{g}(\hat{\mathbf{w}})]^t \mathbf{T} \mathbf{g}(\hat{\mathbf{w}})$$

unter $H_0^w : \mathbf{C} \mathbf{w} = 0$ durch eine $\frac{\chi_f^2}{f}$ -Verteilung approximiert werden, wobei

$$\tilde{f} = \frac{[\text{Spur}(\mathbf{T}[\mathbf{g}'(\hat{\mathbf{w}})] \hat{\mathbf{V}}_N [\mathbf{g}'(\hat{\mathbf{w}})]^t)]^2}{\text{Spur}(\mathbf{T}[\mathbf{g}'(\hat{\mathbf{w}})] \hat{\mathbf{V}}_N [\mathbf{g}'(\mathbf{w})]^t \mathbf{T}[\mathbf{g}'(\hat{\mathbf{w}})] \hat{\mathbf{V}}_N [\mathbf{g}'(\mathbf{w})]^t)}$$

gilt.

Der Beweis dieser Aussagen soll in dieser Arbeit nicht mehr erfolgen, sondern kann als Ausblick auf eine Diplomarbeit oder Dissertation gegeben werden, in welcher die Transformationsmethode auf Faktorpläne oder sogar auf hochdimensionale Daten übertragen werden soll.

Abschließend seien die Ergebnisse zu einer kleinen Niveausimulation mit den transformierten mehrfaktoriellen Verfahren dargestellt.

WTS	Wald-Typ-Statistik
WTS(Logit)	Wald-Typ-Statistik mit Logit transformierten Effekten
WTS(Probit)	Wald-Typ-Statistik mit Probit transformierten Effekten
ATS	ANOVA-TYP-Statistik
ATS(Logit)	ANOVA-Typ-Statistik mit Logit transformierten Effekten
ATS(Probit)	ANOVA-Typ-Statistik mit Probit transformierten Effekten

Abbildung 8.1: Abkürzungen der untersuchten Statistiken

Da diese kleine Simulationsstudie als Ausblick dient, wird die Vorgehensweise nur sehr knapp dargestellt. In dieser wurden unter Hypothese abhängige gleichverteilte Zufallsvariablen erzeugt und diese schließlich durch Verwendung des dritten Strahlensatzes dichotomisiert. Der dritte Strahlensatz ist für die Übertragung der Korrelationsstruktur der Gleichverteilten Zufallsvariablen auf die dichotomen Daten notwendig. Simuliert wurde zum Niveau $\alpha = 0.1, 0.05, 0.01$.

Tabelle 8.1: Niveausimulationen mit 4 Methoden zum Vergleich von Erfolgswahrscheinlichkeiten und Compound-Symmetric-Kovarianzstruktur (n=50,100)

d	Korrelation	n	WTS	WTS(Logit)	WTS(Probit)	ATS	ATS(Logit)	ATS(Probit)
4	0	50	0.8727	0.893	0.8878	0.8977	0.9055	0.9041
			0.9273	0.9453	0.9417	0.9479	0.9521	0.9509
			0.9791	0.9895	0.9873	0.9905	0.9932	0.9921
		100	0.8874	0.8968	0.8953	0.8989	0.9037	0.9019
			0.939	0.9479	0.9465	0.9507	0.9543	0.9532
			0.9847	0.9887	0.9875	0.9883	0.9892	0.9888
	0.3	50	0.878	0.8876	0.8842	0.9014	0.9105	0.9066
			0.9342	0.9414	0.937	0.9527	0.9588	0.9561
			0.9829	0.9875	0.9862	0.9905	0.9943	0.9932
		100	0.8837	0.8887	0.8878	0.8944	0.8995	0.8976
			0.9364	0.9419	0.9407	0.9479	0.9517	0.95
			0.9854	0.9882	0.9873	0.9903	0.9912	0.9909
0.8	50	0.8683	0.8739	0.8703	0.8979	0.902	0.9005	
		0.9339	0.9374	0.9344	0.9528	0.9542	0.9527	
		0.9904	0.9911	0.9904	0.993	0.9945	0.9941	
	100	0.8913	0.8932	0.8917	0.9061	0.9087	0.9081	
		0.9437	0.9446	0.9431	0.954	0.9559	0.9553	
		0.988	0.9883	0.9882	0.9913	0.9922	0.9916	

Tabelle 8.2: Niveausimulationen mit 8 Methoden zum Vergleich der Erfolgswahrscheinlichkeiten und Compound-Symmetric-Kovarianzstruktur (n=50,100)

d	Korrelation	n	WTS	WTS(Logit)	WTS(Probit)	ATS	ATS(Logit)	ATS(Probit)
8	0	50	0.802	0.8495	0.8392	0.906	0.9134	0.9101
			0.8741	0.9115	0.9035	0.9565	0.9593	0.9579
			0.9529	0.9755	0.9714	0.9929	0.9932	0.9927
		100	0.8506	0.8719	0.8679	0.8981	0.902	0.9016
			0.9157	0.9333	0.9295	0.9519	0.9525	0.9524
			0.9761	0.9844	0.9827	0.9896	0.9906	0.9902
	0.3	50	0.8059	0.812	0.8063	0.9147	0.9201	0.9165
			0.8844	0.8855	0.8796	0.9607	0.9635	0.9622
			0.9636	0.9626	0.9609	0.9929	0.9945	0.9938
		100	0.864	0.8665	0.8643	0.9064	0.9085	0.9067
			0.9214	0.9245	0.9221	0.9543	0.9532	0.9549
			0.9787	0.9812	0.9793	0.9908	0.9911	0.9905
0.8	50	0.8305	0.8282	0.8225	0.9236	0.9273	0.9249	
		0.9104	0.9042	0.9016	0.9663	0.9683	0.967	
		0.984	0.981	0.9796	0.9949	0.9961	0.9955	
	100	0.8622	0.8627	0.8595	0.9076	0.9093	0.9078	
		0.9226	0.9234	0.9202	0.9579	0.9582	0.9578	
		0.9835	0.9825	0.9819	0.9929	0.9937	0.9933	

A Grenzwertsätze

Satz A.0.3 (MITTELWERTSATZ)

Die Abbildung $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ sei differenzierbar. Ist $\frac{d}{dx}f(\mathbf{x})$ stetig im Raum $\{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0| < r\}$, dann gilt für $|\mathbf{t}| < r$:

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{t}) = \mathbf{f}(\mathbf{x}_0) + \int_0^1 \mathbf{f}'(\mathbf{x}_0 + u\mathbf{t})d u.$$

Beweis:

Definiere $\mathbf{h}(u) = \mathbf{f}(\mathbf{x}_0 + u\mathbf{t})$. Dann gilt mit der Kettenregel: $\mathbf{h}'(u) = \mathbf{f}'(\mathbf{x}_0 + u\mathbf{t})\mathbf{t}$. Folglich gilt:

$$\int_0^1 \mathbf{f}'(\mathbf{x}_0 + u\mathbf{t})\mathbf{t}d u = \int_0^1 \mathbf{h}'(u)d u = \mathbf{h}(1) - \mathbf{h}(0) = \mathbf{f}(\mathbf{x}_0 + \mathbf{t}) - \mathbf{f}(\mathbf{x}_0).$$

□

Satz A.0.4 (TAYLOR'S THEOREM)

Die Abbildung $f : \mathbb{R}^d \rightarrow \mathbb{R}$ sei zweimal auf $\{\mathbf{x} : |\mathbf{x} - \mathbf{x}_0| < r\}$ stetig differenzierbar, dann gilt für $|\mathbf{t}| < r$:

$$f(\mathbf{x}_0 + \mathbf{t}) = f(\mathbf{x}_0) + f'(\mathbf{x}_0)\mathbf{t} + \mathbf{t}' \int_0^1 \int_0^1 v f''(\mathbf{x}_0 + uv\mathbf{t})d u d v \mathbf{t}.$$

Beweis:

Der Beweis kann im Lehrbuch von FERGUSON [2], Kapitel 7, Theorem 7.4, nachgelesen werden.

□

Satz A.0.5 (LEVY'SCHER GRENZWERTSATZ)

Die Zufallsvariablen $X_i, i = 1, \dots, n$, seien unabhängig und identisch verteilt nach $F(x)$ mit $E(X_1) = \mu$ und $\text{Var}(X_1) = \sigma^2 < \infty$. Dann gilt:

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}x^2}.$$

Beweis:

Der Beweis kann im Vorlesungsskript von Brunner [8] nachgelesen werden.

□

Satz A.0.6 (ZENTRALER GRENZWERTSATZ)

Es seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängig und identisch verteilte Zufallsvektoren mit Erwartungswert μ und Kovarianzmatrix Σ . Dann gilt:

$$\sqrt{n}(\bar{\mathbf{X}}_n - \mu) \xrightarrow{L} \mathcal{N}(\mathbf{0}, \Sigma).$$

Beweis:

Der Beweis kann im Lehrbuch von FERGUSON [2], Kapitel 5, Theorem 5.1, nachgelesen werden. □

Satz A.0.7 (δ -Satz)

Es sei $\phi \in \mathbb{R}$ eine Konstante und T_n eine Folge von Zufallsvariablen, so dass für $n \rightarrow \infty$ gilt $\sqrt{n}(T_n - \phi) \xrightarrow{L} T$. Ferner sei die Funktion $g(t)$ an der Stelle ϕ stetig differenzierbar. Dann gilt

$$\sqrt{n}[g(T_n) - g(\phi)] \xrightarrow{L} g'(\phi) \cdot T.$$

Beweis:

Der Beweis kann im Vorlesungsskript von Brunner [8] nachgelesen werden. □

Satz A.0.8 (CRAMER)

Die Abbildung $\mathbf{g}: \mathbb{R}^d \rightarrow \mathbb{R}^k$ sei auf einer Umgebung von $\mu \in \mathbb{R}^d$ stetig differenzierbar. Es sei \mathbf{X}_n eine Folge von d -dimensionalen Zufallsvektoren mit $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{L} \mathbf{X}$. Dann gilt $\sqrt{n}[\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\mu)] \xrightarrow{L} \mathbf{g}'(\mu) \cdot \mathbf{X}$. Insbesondere gilt, falls $\sqrt{n}(\mathbf{X}_n - \mu) \xrightarrow{L} \mathcal{N}(\mathbf{0}, \Sigma)$,

$$\sqrt{n}(\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\mu)) \xrightarrow{L} \mathcal{N}(\mathbf{0}, \mathbf{g}'(\mu)\Sigma[\mathbf{g}'(\mu)]^t).$$

Beweis:

Der Beweis kann im Lehrbuch von FERGUSON [2], Kapitel 5, Theorem 7.6, nachgelesen werden. □

Satz A.0.9 (Slutzky)

- (1) Sei $\mathbf{X}_n \in \mathbb{R}^k, n \geq 1$, eine Folge von Zufallsvektoren mit $\mathbf{X}_n \xrightarrow{P} \mathbf{a}$, wobei $\mathbf{a} \in \mathbb{R}^k$ konstant ist und sei ferner $g(\cdot)$ stetig in \mathbf{a} , dann gilt:

$$g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{a}).$$

- (2) Sei $\mathbf{X}_n \in \mathbb{R}^k, n \geq 1$, eine Folge von Zufallsvektoren mit $\mathbf{X}_n \xrightarrow{L} \mathbf{X} \sim F(\mathbf{x})$ und sei $g(\cdot)$ eine F -f.ü stetige Funktion. Dann gilt

$$g(\mathbf{X}_n) \xrightarrow{L} g(\mathbf{X}) \sim F_g(\mathbf{x}).$$

Beweis:

Der Beweis kann im Vorlesungsskript von Brunner [8] nachgelesen werden. □

B Software

B.1 Zweistichprobenproblem-verbundene Stichproben

Dargestellt wird exemplarisch ein Programm, welches die Konfidenzgrenzen der Intervalle des verbundenen Zweistichprobenproblems und die mathematischen a-priori-Maße berechnet.

```
proc iml worksize=120 ;
reset nolog;
start ki(n,alfa,nscan,ko,vopt,aopt,theta) ;

*---erzeuge die Integrationsvariablen-----* ;
vp1=j(nscan-1,1,0) ;
do k=1 to nscan-1 ;
vp1[k]=k/nscan ;
end ;
vp2=vp1 ;
nvp1 = nrow(vp1) ;
nvp2 = nrow(vp2) ;
nstat = nrow(col) - 1 ;
nstat1 = nrow(colcp) ;
u = PROBIT(1-alfa/2) ;
kon = u / sqrt(N) ;
konst= PROBIT(1-alfa/2) ;
waoben=j(n+1,n+1,0) ;
waunten=j(n+1,n+1,0) ;
agoben=j(n+1,n+1,0) ;
agunten=j(n+1,n+1,0) ;
grenzeo=j(n+1,n+1,0) ;
grenzeu=j(n+1,n+1,0) ;
tabelleu=j(n+1,n+1,0) ;
tabelleo=j(n+1,n+1,0) ;
newoben=j(n+1,n+1,0) ;
newunten=j(n+1,n+1,0) ;
erg=j(n+1,n+1,0) ;
coverage=j(nvp1,nvp2,0) ;
coverageag=j(nvp1,nvp2,0) ;
coveragetraf=j(nvp1,nvp2,0) ;
coveragenew=j(nvp1,nvp2,0) ;
längewald=j(nvp1,nvp2,0) ;
längeagresti=j(nvp1,nvp2,0) ;
längegetraf=j(nvp1,nvp2,0) ;
länge=new=j(nvp1,nvp2,0) ;
vk1 = j(n+1,1,0) ;
vk2 = j(n+1,1,0) ;
nuk2 = j(n+1,1,0) ;
multi=j(n+1,n+1,0) ;
nenner3=j(n+1,n+1,0) ;
*-----Beginne die Scanning-Schleife-----* ;
DO y=1 TO nvp1 ;
DO z=1 TO nvp2 ;
p1=vp1[y] ;
p2=vp2[z] ;
*-----Übergebe die Einträge aus dem Odds-Ratio-----* ;
if theta=1 then do ;
p11=p1*p2 ;
end ;
if theta >1 then do ;
A=(p2*(theta-1)+p1*(theta-1)+1)/(2*(theta-1)) ;
p11=A-sqrt(A**2- (theta*p2*p1)/(theta-1)) ;
```

```

end ;
if theta<1 then do ;
A=(p2*(theta-1)+p1*(theta-1)+1)/(2*(theta-1)) ;
p11=A+sqrt(A**2-(theta*p2*p1)/(theta-1)) ;
end ;
*--Erzeuge die hypothetischen Wahrscheinlichkeiten-----* ;
pc=p1-p11 ;
pb=p2-p11 ;
delta=pb-pc ;
*-----Berechne Matrix der multinomialkoeffizienten-----* ;
nenner1=1 ;
nenner2=1 ;
nenner3=1 ;
DO i=0 TO n ;
    k1 = i-1 ;
    nenner1=nenner1*i ;
if nenner1=0 then do ;
    nenner1=1 ;
end ;
DO j=0 TO n-i ;
    k2= j-1 ;
    zähler=1 ;
    do x=1 TO n ;
        zähler=zähler*x ;
    end ;
    nenner2=nenner2*j ;
    if nenner2=0 then do ;
        nenner2=1 ;
    end ;
    DO blub=0 TO n-i-j ;
        nenner3=nenner3*blub ;
        if nenner3 =0 then do ;
            nenner3=1 ;
        end ;
        end ;
        quotient=zähler /(nenner2*nenner1*nenner3) ;
        multi[i+1,j+1]=quotient*(pc+delta)**(j)*pc**(i) *
        (1-2*pc-delta)**(n-i-j) ;
*-----Berechne nun die Intervalle-----* ;

*-----Wald-Konfidenzintervalle-----* ;
pdach=1/n*(i-j) ;
var1=1/n*(i+j) ;
var2=(i/n-j/n)**2 ;
varianz=(var1-var2) ;
waunten[i+1,j+1]=pdach-kon*sqrt(varianz) ;
waoben[i+1,j+1]=pdach+kon*sqrt(varianz) ;
*-----Agresti-Konfidenzintervall-----* ;
zu=2 ;
nstar=n+zu ;
konstar=u /sqrt(nstar) ;
istar=i+zu/4 ;
jstar=j+zu/4 ;
pdachag=1/nstar*(istar-jstar) ;
varlag=(istar/nstar+jstar/nstar) ;
var2ag=(istar/nstar-jstar/nstar)**2 ;
varianzag=(varlag-var2ag) ;
agunten[i+1,j+1]=pdachag-konstar*sqrt(varianzag) ;
agoben[i+1,j+1]=pdachag+konstar*sqrt(varianzag) ;
*-----Transformiertes Intervall-----* ;
p21=i/n ;
p12=j/n ;
pd= p21-p12 ;
var1dach=p21*(1-p21) ;
var2dach=p12*(1-p12) ;
var3dach=2*p12 * p21 ;
varianztra=var1dach+var2dach+var3dach+vopt/(n**aopt) ;
stand=sqrt(varianztra) ;
fak=(1+ko)/((1+ko)**2-(p21 - p12)**2) ;

```

```

zählertra=1+ko+pd ;
nennertra=1+ko-pd ;
trafo=1/2*log(zählertra/nennertra) ;
tabelleu[i+1,j+1]=trafo-kon*fak*stand ;
tabelleo[i+1,j+1]=trafo+kon*fak*stand ;
*grenze mit rücktrafo;
grenzeu[i+1,j+1]=(exp(2*tabelleu[i+1,j+1])+
ko*(exp(2*tabelleu[i+1,j+1])-1)-1) /
(exp(2*tabelleu[i+1,j+1])+1) ;
grenzeo[i+1,j+1]=(exp(2*tabelleo[i+1,j+1])+
ko*(exp(2*tabelleo[i+1,j+1])-1)-1) /
(exp(2*tabelleo[i+1,j+1])+1) ;
end ;
end ;

*-----passe die Hybrid Grenzen an-----*;
newobenja=(newoben<1) ;
newuntenja=(newunten>-1) ;
newoben=(newoben#newobenja)+(newobenja=0) ;
newunten=(newunten#newuntenja)-(newuntenja=0) ;
länew=(newoben-newunten)#multi ;
*-----Berechne nun den Indikator für hybrid-----*;
erg11new=(newoben - (p1-p2)>0) ;
erg12new=((p1-p2) - newunten >0) ;
erg1new=((erg11new+erg12new)=2) ;
*---Berechne Coverage Probability für hybrid-----*;
ergnew=erg1new#multi ;
coveragenew[y,z]=sum(ergnew) ;
längenew[y,z]=sum(länew) ;
*-----Passe die Wald-Grenzen an-----*;
waobenja=(waoben<1) ;
wauntenja=(waunten>-1) ;
waoben=(waoben#waobenja)+(waobenja=0) ;
waunten=(waunten#wauntenja)-(wauntenja=0) ;
längewa=(waoben-waunten)#multi ;
*-----Berechne nun den Indikator für wald-----*;
erg11=(waoben - (p1-p2)>0) ;
erg12=((p1-p2) - waunten >0) ;
erg1=((erg11+erg12)=2) ;
*-----Berechne Coverage Probability für wald-----*;
erg=erg1#multi ;
coverage[y,z]=sum(erg) ;
längewald[y,z]=sum(längewa) ;
*-----Agesti-Grenzen-----*;
agobenja=(agoben<1) ;
aguntenja=(agunten>-1) ;
agoben=(agoben#agobenja)+(agobenja=0) ;
agunten=(agunten#aguntenja)-(aguntenja=0) ;
längeag=(agoben-agunten)#multi ;
*-----Berechne nun den Indikator für Agresti-----*;
erg11ag=(agoben - (p1-p2)>0) ;
erg12ag=((p1-p2) - agunten >0) ;
erg1ag=((erg11ag+erg12ag)=2) ;

*---Berechne Coverage Probability für Agresti-----*;
ergag=erg1ag#multi ;
coverageag[y,z]=sum(ergag) ;
längeagresti[y,z]=sum(längeag) ;
*-----Transformierte-Grenzen-----*;
trobenja=(grenzeo<1) ;
truntenja=(grenzeu>-1) ;
grenzeo=(grenzeo#trobenja)+(trobenja=0) ;
grenzeu=(grenzeu#truntenja)-(truntenja=0) ;
längetr=(grenzeo-grenzeu)#multi ;
*-----Berechne nun den Indikator für Transformiert-----*;
erg11tr=(grenzeo - (p1-p2)>0) ;
erg12tr=((p1-p2) - grenzeu >0) ;
erg1tr=((erg11tr+erg12tr)=2) ;
*---Berechne Coverage Probability für Transformiert---*;

```

```

ergtr=ergltr#multi ;
coveragetraf[y,z]=sum(ergtr) ;
längetraf[y,z]=sum(längetr) ;
end ;
end ;
*-----Maße für das Wald-Intervall-----*;
mittelwa=sum(coverage)/(nvpl*nvp2) ;
lwa=sum(längewald)/(nvpl*nvp2) ;
mittelwakonserva=sum((coverage>=0.95)*(coverage-0.95))
/(nvpl*nvp2) ;
mittelwaantikonserva=sum((coverage<=0.95)*
(0.95-coverage))/(nvpl*nvp2) ;
mittelwafehler=sum(abs(coverage-0.95))/(nvpl*nvp2) ;
mittelwastreuung=sqrt(sum( (Coverage-0.95)**2)
/(nvpl*nvp2)) ;
*---Maße für Agresti---*;
mittelag=sum(coverageag)/(nvpl*nvp2) ;
lag=sum(längeagresti)/(nvpl*nvp2) ;
mittelagkonserva=sum((coverageag>=0.95)#
(coverageag-0.95))/(nvpl*nvp2) ;
agantikonserva=(coverageag<=0.95);
mittelagantikonserva=sum((coverageag<=0.95)#
(0.95-coverageag))/(nvpl*nvp2) ;
mittelagfehler=sum(abs(coverageag-0.95))/(nvpl*nvp2) ;
mittelagstreuung=sqrt(sum( (Coverageag-0.95)**2)/
(nvpl*nvp2)) ;
*-----Maße für Transformiert-----*;
mitteltraf=sum(coveragetraf)/(nvpl*nvp2) ;
ltraf=sum(längetraf)/(nvpl*nvp2) ;
mitteltrafkonserva=sum((coveragetraf>=0.95)#
(coveragetraf-0.95))/(nvpl*nvp2) ;
untertraf=(coveragetraf<=0.95) ;
antikonservatraf=untertraf#((0.95)-coveragetraf) ;
mitteltrafantikonserva= sum(antikonservatraf)/(nvpl*nvp2);
mitteltraffehler=sum(abs(coveragetraf-0.95))/(nvpl*nvp2);
mitrafstreuung=sqrt(sum( (Coveragetraf-0.95)**2)/
(nvpl*nvp2)) ;
lnew=sum(längeneu)/(nvpl*nvp2) ;
print mittelwa lwa mittelwakonserva mittelwaantikonserva
mittelwafehler mittelwastreuung ;
print mittelag lag mittelagkonserva mittelagantikonserva
mittelagfehler mittelagstreuung;
print mitteltraf ltraf mitteltrafkonserva
mitteltrafantikonserva
mitteltraffehler mitrafstreuung ;
finish;
run ki( 30,0.05,5,1/50,0.33,1,1 ) ;
quit ;

```

Literaturverzeichnis

- [1] KRENGEL, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Braunschweig, Wiesbaden, sechste Auflage.
- [2] FERGUSON, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall, London, erste Auflage.
- [3] ROHATGI, V. K. (1984). *Statistical Inference*. John Wiley & Sons, New York.
- [4] AGRESTI, A. (1984). *Analysis of ordinal categorical data*. John Wiley & Sons, New York.
- [5] SCHMETTERER, L. (1966). *Mathematische Statistik*. Springer, Wien, New York, zweite Auflage.
- [6] WERNER, J. (1992). *Numerische Mathematik I*. Vieweg, Braunschweig, Wiesbaden.
- [7] WERNER, J. (1984). *Optimization Theory and Applications*. Vieweg, Braunschweig, Wiesbaden.
- [8] BRUNNER, E. (2004). *Angewandte Statistik I*. Vorlesungsskript, Universität Göttingen.
- [9] BRUNNER, E. (2005). *Angewandte Statistik 2*. Vorlesungsskript, Universität Göttingen.
- [10] BRUNNER, E. (2004). *Theory of rank tests in factorial designs*. Vorlesungsskript, Universität Göttingen.
- [11] XIAO-HUA,Z. AND CHUNMING,L.*Improving interval estimation of binomial proportions*.
- [12] CHRISTOPHLIEMK, O. (2001). *Die Transformationsmethode im gemischten nichtparametrischen Modell mit Kovariablen*. Diplomarbeit, Institut für Mathematische Stochastik Göttingen.
- [13] DOMHOF, S.(2001). *Nichtparametrische relative Effekte*. Dissertation, Universität Göttingen, Göttingen.
- [14] NEWCOMBE, R. (1998). Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* **17**, 2635-2650.
- [15] VOS, P. AND HUDSON,S. (2005). Evaluation Criteria for Discrete Confidence Intervals: Beyond Coverage and Length. *American Statistical Association* **59**, 137-142.
- [16] ROCKE, D.M. (1993). On the Beta Transformation Family. *American Statistical Association* **35**, 72-81.
- [17] AGRESTI, A. AND COULL, B.A. (1998). Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions. *The American Statistician* **52**, 119-126.
- [18] AGRESTI, A. (1996). An Introduction to Categorical Data Analysis. *New York: Wiley*.
- [19] AGRESTI, A. AND CAFFO, B. (2000). Simple and Effectice Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two failures. *The American Statistician* **54**, 281-287.
- [20] AGRESTI, A. AND MIN, Y. (2005). Simple improved confidence intervals for comparing matched proportions. *Statistics in Medicine* **24**, 729-740.

-
- [21] BOYSEN, L. (2002). *Analyse intra-individuellen Effekten bei longitudinalen Daten*. Diplomarbeit, Institut für Mathematische Stochastik Göttingen.
- [22] BOX, G.E.P (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems I. Effect of the Inequality of Variance in the One-Way-Classification. *The Annals of Mathematical Statistics* **25**, 290-302
- [23] BRUNNER, E., DETTE, H., MUNK, A. (1997). Box-Type Approximations in Nonparametric Factorial Designs. *Journal of the American Statistical Association* **92**, 1497-1502.
- [24] BRUNNER, E., DOMHOF, S., LANGER, F. (2001). Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Wiley, New York* .
- [25] BRUNNER, E., MUNZEL, U., PURI, M. L. (1999). Rank-Score Tests in Factorial Designs with Repeated Measures. *Journal of Multivariate Analysis* **70**, 286-317.
- [26] TANGO, T. (1998) . Equivalence Test and Confidence Interval for the Difference in Proportions for the Paired-Sample Design. *Statistics in Medicine* **17**, 891-908.
- [27] TANGO, T. (1999) . Letter to the Editor. *Statistics in Medicine* **18**, 3511-3513.
- [28] TANG, MAN-LAI, TANG, NIAN-SHENG, CHAN, IVAN S.F. . Confidence interval construction for proportion difference in small-sample paired studies. *Statistics in Medicine* **24**, 3565-3579.
- [29] WERNER, C. (2006). *Nichtparametrische Analyse von diagnostischen Tests*. Dissertation, Universität Göttingen, Göttingen.
- [30] NEWCOMBE, R. G. . Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* **17**, 873-890.

Danksagung:

Ganz herzlich möchte ich mich bei Herrn Prof. Dr. E. Brunner für die hervorragende Betreuung der Arbeit und die sehr guten Arbeitsmöglichkeiten in der Abteilung Medizinische Statistik der Universität Göttingen bedanken. Ganz besonders bedanke ich mich bei ihm für die Vergabe dieses sehr aktuellen Forschungsthemas und freue mich sehr, dass ich meine Ergebnisse auf Kongressen vorstellen darf. Zu besonderem Dank bin ich Frau Prof. Dr. Schöbel der Universität Göttingen für die Beratung bei den Optimierungsalgorithmen verpflichtet.

Ohne die sehr gute EDV-Ausstattung der Abteilung für Medizinische Statistik wäre diese Arbeit in der Form nie entstanden. Die bivariaten numerischen Integrationen sind in der dargebrachten Form im Programmpaket 'R' nicht durchführbar, sondern entstanden schließlich nach Umprogrammierung in SAS-IML. Weiter danke ich für die Geduld der Mitarbeiter, da die Optimierungsalgorithmen doch fast alle Rechner des CIP-Raumes so manche Woche in Anspruch nahmen.

Außerdem danke ich allen Mitarbeitern und Kollegen für die tollen Kuchen und allerlei Ratschläge und Tipps! Vor allem bei anfänglichen Programmierstolpern im doch nicht so ganz leichten SAS-IML.

Last but not least danke ich meiner meiner Familie und natürlich allen Freunden. Hervorgehoben seien jetzt die, die diese Arbeit korrekturlesen.