

# Schwierigkeiten mit Konfidenz-Intervallen

JÖRG MEYER, HAMELN

**Zusammenfassung:** In diesem Aufsatz geht es um die Berechnung von Konfidenz-Intervallen für Binomialverteilungen, die Aussagen über die Einzelerfolgs-Wahrscheinlichkeit  $p$  machen.

Nun kann man Konfidenz-Intervalle unterschiedlich konstruieren und deren Eigenschaften mit der heute verfügbaren Software auch jeweils bequem visualisieren. Es ist für Lehrende günstig, hier über erweitertes Hintergrundwissen zu verfügen. Dieses wird in diesem Aufsatz geleistet: Es werden mehrere verschiedene Arten vorgestellt, zweiseitige Konfidenz-Intervalle für  $p$  zu konstruieren (es handelt sich um die Konfidenz-Intervalle nach WILSON, WALD und CLOPPER/PEARSON), und es werden ihre Vor- und Nachteile beschrieben:

Dies betrifft einerseits konzeptuelle Probleme (vor allem beim WALD-Intervall), andererseits aber auch die Tatsache, dass die messbare Überdeckungs-Häufigkeit von der vorgegebenen Sicherheits-Wahrscheinlichkeit deutlich abweichen kann: Die Überdeckungs-Häufigkeit ist i. a. bei CLOPPER/PEARSON zu groß und bei WALD zu klein, während bei WILSON die Abweichungen auch für kleine Stichproben gering sind.

## Um was es geht

Wir betrachten eine Binomialverteilung mit unbekannter Einzelerfolgs-Wahrscheinlichkeit  $p$ . Eine  $n$ -elementige Stichprobe führe zu  $H$  Erfolgen. Was kann man dann über  $p$  aussagen? Konfidenz-Intervalle überdecken  $p$  mit einer (vorab festzulegenden) Sicherheits-Wahrscheinlichkeit  $\delta$ .

Die Abbildungen dieses Artikels wurden (bis auf Abb. 9) mit GeoGebra 4.2 erstellt; vgl. Meyer (2011). Die zugehörigen Dateien können vom Autor angefordert werden.

## 1 Prognose-Intervalle

Da es bei Konfidenz-Intervallen um den Schluss von einer Stichprobe auf die Grundgesamtheit geht, ist es sinnvoll, zunächst den umgekehrten Schluss zu thematisieren: Welche Aussagen lassen sich über Stichproben bei bekannter Grundgesamtheit machen (vgl. Lergenmüller/Schmidt/Krüger (2012, Kap. 1.3 und 5.1) oder Meyer (2011)), wenn man die Binomialverteilung durch die Normalverteilung approximiert? Dies führt zum *Prognose-Intervall*: Man gebe eine

Sicherheits-Wahrscheinlichkeit  $\delta$  vor und bestimme  $k$  nach der Beziehung

$$\delta = \int_{-k}^k \varphi(x) \cdot dx$$

(mit  $\varphi$  als standardisierter Normalverteilung).

Ist die Grundgesamtheit binomialverteilt (mit Einzelerfolgs-Wahrscheinlichkeit  $p$ ) und bezeichnet  $H$  die Anzahl der Erfolge in einer  $n$ -elementigen Stichprobe, so gilt mit Wahrscheinlichkeit  $\delta$  die Doppel-Ungleichung

$$|H - \mu| \leq k \cdot \sigma \quad (1)$$

bzw.

$$|H - n \cdot p| \leq k \cdot \sqrt{n \cdot p \cdot (1 - p)}$$

(„ $\sigma$ -Regel“). Nach Division durch  $n$  und nach Übergang zur relativen Häufigkeit  $h = \frac{H}{n}$  bestimmen sich die Grenzen des Prognose-Intervalls aus

$$h = p \pm k \cdot \sqrt{\frac{p \cdot (1 - p)}{n}}. \quad (2)$$

Die „Faustregel“  $\sigma > 3$  sollte erfüllt sein.

*Beispiel:* Man hat eine gezinkte Münze, die mit 40 %-iger Wahrscheinlichkeit auf die nationale Seite fällt.  $H$  zählt die Anzahl der so definierten Erfolge. Man wirft die Münze 100-mal. Dann liegt die Anzahl  $H$  der Erfolge mit Wahrscheinlichkeit 95 % zwischen 30 und 50.

## 2 Konfidenz-Intervalle nach WILSON

### 2.1 Vorgehensweise

Nun zur *Umkehrung*: Man hat eine vermutlich gezinkte Münze. Die Wahrscheinlichkeit  $p$ , mit der sie auf die nationale Seite fällt, ist unbekannt.

Man wirft 100-mal und hat 40 Erfolge. Was kann man über  $p$  sagen?

Um eine Fokussierung auf die Wahrscheinlichkeit eines Einzel-Ereignisses („ $p$  so, dass  $\text{prob}(H=40)$  maximal“) zu vermeiden, ist es der wohl beste Weg, mit fiktiven Werten für  $p$  und den zugehörigen Prognose-Intervallen zu arbeiten, d. h.: In (1) ist nicht  $p$  bekannt, sondern  $H$ .

Da die 95 %-Prognose-Intervalle zu  $p = 31\%$  und zu  $p = 50\%$  gerade noch  $H$  enthalten, ist [31 %; 50 %] das 95 %-Konfidenz-Intervall (für den unbekannt Parameter  $p$ .) Das so erhaltene Konfidenz-Intervall wird in der Literatur nach EDWIN B. WILSON (1927) benannt (in der englischen Literatur auch „score interval“).

Das WILSON-Intervall besteht somit aus den Parametern passender fiktiver Prognose-Intervalle, wenn man die Binomialverteilung durch die Normalverteilung annähert.

## 2.2 Eine erste Schüler-Schwierigkeit

„Also ist mit 95 %-iger Wahrscheinlichkeit die Münze gezinkt“. Dies ist natürlich ein sehr vorschneller Schluss, denn es geht gar nicht darum, *ob* die Münze gezinkt ist oder nicht, sondern *wie* sie gezinkt ist.

Diese Schüler-Schwierigkeit ist nicht für das WILSON-Intervall typisch, sondern betrifft alle Arten von Konfidenz-Intervallen.

## 2.3 Eine zweite Schüler-Schwierigkeit

Zum letzten Absatz semantisch gleichwertig, aber mit anderer (falscher) Emphase, ist die folgende Formulierung:

„Die gesuchte Einzelerfolgs-Wahrscheinlichkeit  $p$  liegt mit Wahrscheinlichkeit 95 % im WILSON-Intervall“.

In Wahrheit ist aber nicht (das feststehende, jedoch unbekannt)  $p$  die Zufallsgröße, sondern es sind die Grenzen des Konfidenz-Intervalls.

Auch diese Schüler-Schwierigkeit ist nicht für das WILSON-Intervall typisch, sondern betrifft alle Arten von Konfidenz-Intervallen.

Der zweiten Schwierigkeit wird man wohl am besten durch ein Diagramm wie in Abb. 1 abhelfen, das viele (auf der von  $p$  abhängigen Zufallsgröße  $H$  beruhenden) Konfidenz-Intervalle zeigt (und diese auch auf Knopfdruck neu entstehen lässt).

Eine Simulation ergab: Ist  $\delta = 0,95$  sowie  $n = 25$  und  $p = 0,4$ , so überdecken 93 von 100 Konfidenz-Intervallen den wahren Wert 0,4. Wiederholungen dieser Simulation ergaben vergleichbare Werte.

## 2.4 Berechnung der WILSON-Intervalle

Man muss die Doppel-Ungleichung

$$|h - p| \leq k \cdot \sqrt{\frac{p \cdot (1 - p)}{n}} \quad (2)$$

nach  $p$  auflösen. Das ist mit graphischen Taschenrechnern (oder noch leistungsfähigeren Werkzeugen) problemlos möglich.

Will man (2) „zu Fuß“ behandeln, ist eine quadratische Gleichung zu lösen, deren Lösungen die Grenzen des Konfidenz-Intervalls sind. Das Ergebnis ist

$$p = \frac{h \cdot n + \frac{k^2}{2} \pm k \cdot \sqrt{n \cdot h \cdot (1 - h) + \frac{k^2}{4}}}{n + k^2} = \frac{h + \frac{k^2}{2 \cdot n} \pm \frac{k}{\sqrt{n}} \cdot \sqrt{h \cdot (1 - h) + \frac{k^2}{4 \cdot n}}}{1 + \frac{k^2}{n}} \quad (3)$$

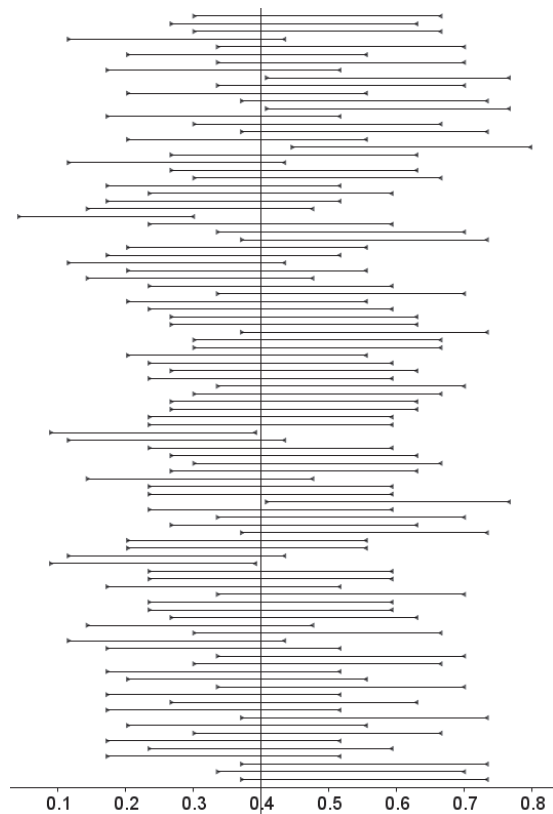


Abb. 1: WILSON-Konfidenz-Intervalle zu  $p = 0,4$

## 2.5 Liegt das WILSON-Konfidenz-Intervall immer in [0; 1]?

Diese Frage ist mit „ja“ zu beantworten wegen

$$h \cdot n + \frac{k^2}{2} \geq k \cdot \sqrt{n \cdot h \cdot (1 - h) + \frac{k^2}{4}}$$

$\Leftrightarrow$

$$h^2 \cdot n^2 + h \cdot n \cdot k^2 \geq k^2 \cdot n \cdot h \cdot (1 - h)$$

und wegen

$$h \cdot n + \frac{k^2}{2} \pm k \cdot \sqrt{n \cdot h \cdot (1 - h) + \frac{k^2}{4}} \leq n + k^2$$

$\Leftrightarrow$

$$k^2 \cdot n \cdot (h - 1) \leq n^2 (1 - h).$$

## 2.6 Zur Überdeckungs-Häufigkeit

Aufgrund von Abb. 1 und vor allem aufgrund der Tatsache, dass bei der Berechnung der WILSON-Intervalle die Gleichung (2) nur „umgekehrt betrachtet“ wird, sollte man erwarten, dass das zu  $\delta$  gehörige Konfidenz-Intervall den wahren Parameter  $p$  mit etwa der relativen Häufigkeit  $\delta$  überdeckt – allerdings nur „etwa“, da bei der Konstruktion des WILSON-Intervalls die Approximation der Binomial- durch die Normalverteilung verwendet wurde.

Die Erwartung tritt auch tatsächlich ein, wie man an Abb. 2 sieht oder wie man bei Agresti und Coull (1998) nachlesen kann.

In Abb. 2 ist  $\delta = 0,95$  und  $n = 25$ . Auf der Rechtsachse wird  $p$  aufgetragen; mit  $p$  und  $n$  wurde 1000-mal die binomialverteilte Zufallsgröße  $H$  ermittelt und das zugehörige Konfidenz-Intervall berechnet. Nach oben ist der Prozentsatz derjenigen Konfidenz-Intervalle aufgetragen, die  $p$  enthalten. Dieser Prozentsatz wird hier mit *Überdeckungs-Häufigkeit* bezeichnet.

## 2.7 WILSON mit Stetigkeitskorrektur

Die Doppel-Ungleichung (1) schreibt sich nun als

$$|H - \mu| \leq k \cdot \sigma + \frac{1}{2}$$

und führt zu komplizierteren Grenzen des Konfidenz-Intervalls.

Agresti und Coull (1998, S. 120 und S. 125) empfehlen die Stetigkeits-Korrektur nicht, da die aus ihnen resultierenden Überdeckungs-Häufigkeiten weiter von  $\delta$  entfernt sind als bei den unkorrigierten WILSON-Intervallen.

## 2.8 Fazit

Die WILSON-Konfidenz-Intervalle haben Überdeckungs-Häufigkeiten, die auch für kleine Stichproben in der Nähe der Sicherheits-Wahrscheinlichkeit liegen. Nur für ganz kleine und ganz große Werte von

$p$  sind die Abweichungen nennenswert. Die Ermittlung der WILSON-Intervalle ist mit Technologieeinsatz problemlos geworden.

Die Abweichungen zwischen Überdeckungs-Häufigkeit und Sicherheits-Wahrscheinlichkeit können positiv oder negativ sein. In vielen Fällen ist man nur daran interessiert, dass die Abweichungen betragsmäßig klein sind, und dann bietet sich das WILSON-Intervall an.

## 3 Konfidenz-Intervalle nach WALD

### 3.1 Vorgehensweise

Das Ergebnis (3) in 2.4 vereinfacht sich für große Werte von  $n$ , wenn man  $\frac{k^2}{n}$  (aber nicht  $\frac{k}{\sqrt{n}}$ ) vernachlässigt (Engel 1987, S. 138); man erhält

$$p = h \pm k \cdot \sqrt{\frac{h \cdot (1 - h)}{n}}. \quad (4)$$

Man hätte auch sagen können: In (2) wird das unbekannte  $p$  unter dem Wurzelzeichen durch die bekannte relative Häufigkeit  $h$  ersetzt.

Das nach (4) bestimmte Konfidenz-Intervall wird in der Literatur nach ABRAHAM WALD (1939) benannt.

Die Beziehung (4) liefert ein zu  $h$  *symmetrisches* Konfidenz-Intervall (die Beziehung (3) leistet das nicht).

Der große Vorteil von (4) liegt in der einfachen Berechenbarkeit des Konfidenz-Intervalls; ein Argument, das allerdings bei Technologie-Einsatz seine Überzeugungskraft verliert.

Die WALD-Intervalle existieren nicht für jedes  $H$ , wie etwa das Beispiel  $n = 25$ ,  $\delta = 0,95$  und  $H = 2$  zeigt: Das zugehörige Konfidenz-Intervall nach (4) liegt nicht  $[0; 1]$ .

Ist  $h = 0$  oder  $h = 1$ , so hat das WALD-Intervall die Länge 0, existiert also in einem eigentlichen Sinne ebenfalls nicht.

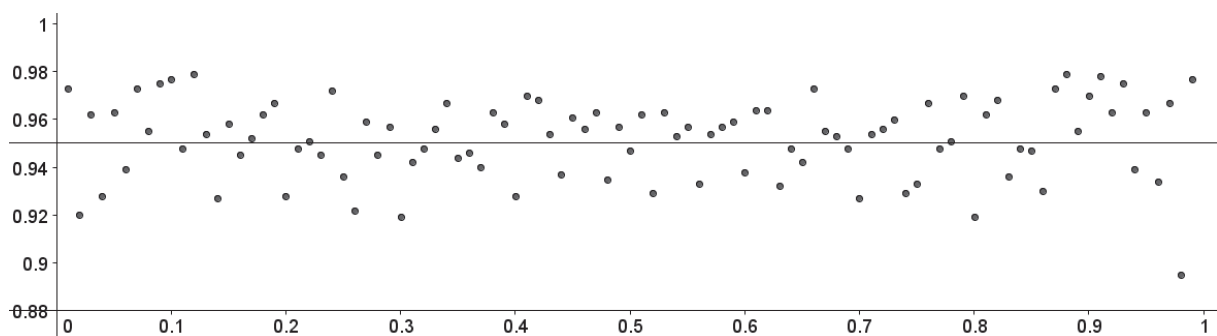


Abb. 2: Überdeckungs-Häufigkeit der WILSON-Konfidenz-Intervalle für  $\delta = 0,95$  und  $n = 25$

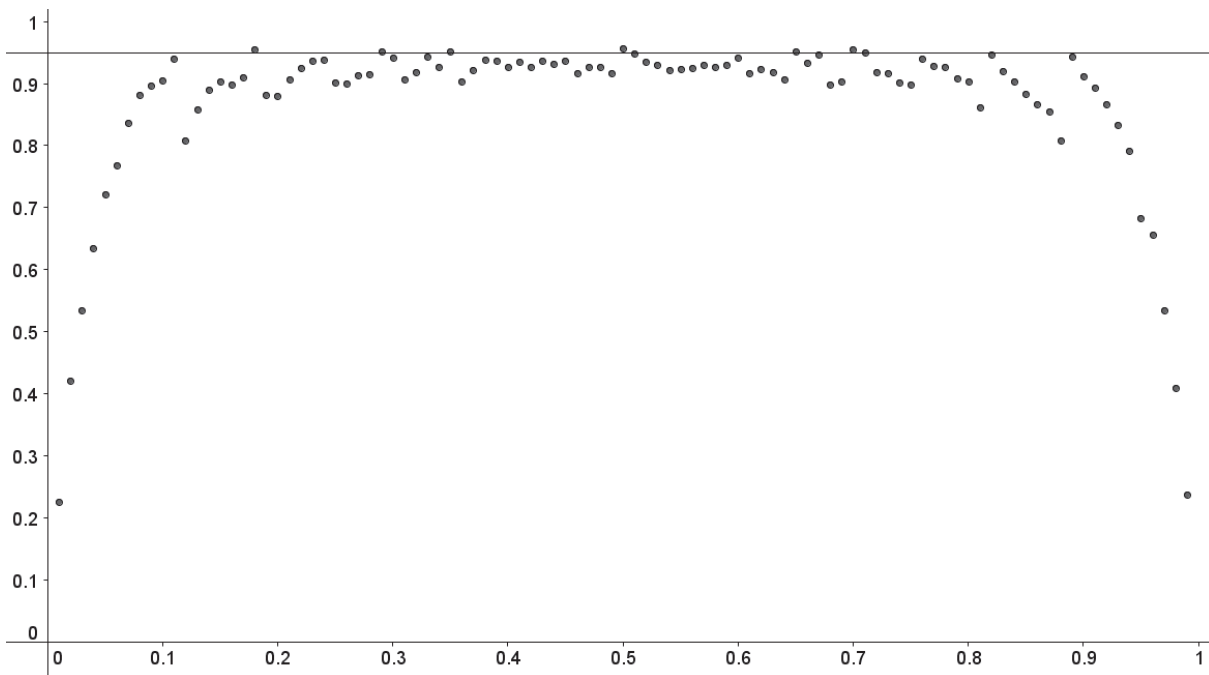


Abb. 3: Überdeckungs-Häufigkeit der WALD-Konfidenz-Intervalle für  $\delta = 0,95$  und  $n = 25$

### 3.2 Eine erste Schüler-Schwierigkeit

Man fragt sich, warum in (2) das unbekannte  $p$  nicht durchgängig an allen Stellen durch  $h$  ersetzt wird. Um diese Frage ordentlich zu beantworten, wird man den oben geschilderten Weg über die Lösung der quadratischen Gleichung gehen müssen.

### 3.3 Eine zweite Schüler-Schwierigkeit

Es gibt ein *strukturelles Problem*: Das angenäherte WALD-Intervall bestimmt sich nach (4) zu

$$p = h \pm k \cdot \sqrt{\frac{h \cdot (1-h)}{n}}.$$

Andererseits ist das Prognose-Intervall für die relative Häufigkeit  $h$  nach (2) gegeben durch

$$h = p \pm k \cdot \sqrt{\frac{p \cdot (1-p)}{n}}.$$

Dies bedeutet:

Beim WALD-Intervall ist der Übergang von  $p$  nach  $h$  *formal derselbe* wie beim Prognose-Intervall der Übergang von  $h$  nach  $p$ !

Schülerinnen und Schüler benutzen daher für Konfidenz-Intervalle und Prognose-Intervalle *denselben* Formalismus! Dies erschwert erheblich die Unterscheidung zwischen beiden Intervallarten, insbesondere bei folgender Aufgabe:

„Letztes Jahr betrug der Anteil der Anhänger für Partei A 30 %. Dieses Jahr hat man bei einer 100-ele-

mentigen Stichprobe 40 Anhänger gemessen.“ Hier kann man ausgehend von den 30 % ein Prognose-Intervall erstellen und nachsehen, ob die 40 darin enthalten sind, oder ausgehend von den 40 „Treffern“ ein Konfidenz-Intervall erstellen und nachsehen, ob es die 30 % überdeckt. Diese beiden Sichtweisen sind kaum noch zu unterscheiden, wenn für beide Wege derselbe Formalismus verwendet wird.

### 3.4 Zur Überdeckungs-Häufigkeit der WALD-Intervalle

Unternimmt man eine zu 2.6 analoge Simulation, bekommt man Abb. 3; dort ist wieder  $\delta = 0,95$  und  $n = 25$ . Man erkennt, dass die Überdeckungs-Häufigkeiten der WALD-Intervalle viel weiter von  $\delta$  entfernt sind, als es beim WILSON-Intervall der Fall ist.

Abb. 4 gehört zu  $\delta = 0,99$  und  $n = 20$ ; sie bestätigt die Figur 4 in Brown et al. (2001; S. 106) und zeigt die starke Oszillation der Überdeckungs-Häufigkeiten. Nebenbei: Für jeden Wert von  $p$  ist die Überdeckungs-Häufigkeit echt kleiner als die Sicherheits-Wahrscheinlichkeit  $\delta$ .

Brown et al. (2001) sowie Agresti und Coull (1998) berichteten über analoge Befunde und *raten daher dringend vom WALD-Intervall ab*.

Es sollte nur für  $n > 100$  und  $0,1 < p < 0,9$  verwendet werden. Man beachte auch die vielen Kommentare, die im Anschluss an Brown et al. (2001) abgedruckt wurden.

### 3.5 Zur Symmetrie der WALD-Intervalle

Dass das WALD-Intervall für kleine und große  $p$  derart kleine Überdeckungs-Häufigkeiten aufweist, lässt sich damit erklären, dass es zu  $h$  symmetrisch ist, obwohl die zugehörigen Binomialverteilungen sehr asymmetrisch sind.

Daher ist die Überdeckungs-Häufigkeit beim WILSON-Intervall i. a. viel näher am Wert der Sicherheits-Wahrscheinlichkeit als beim WALD-Intervall; das gilt auch für die im nächsten Abschnitt behandelten CLOPPER/PEARSON-Intervalle (Agresti und Coull 1998 oder Brown et al. 2001; S. 110 oder Pires und Amado 2008; S. 181 und 183).

Man sollte daher ein Konfidenz-Intervall *nicht* so auffassen, dass es einen „Schätzwert  $\pm$  Fehler“ angibt.

### 3.6 „Konvergenz“ der Überdeckungs-Häufigkeiten?

Abb. 5 illustriert und erweitert Table 1 in Brown et al. (2001; S. 104): Nun wird  $p = 0,5$  festgelassen und

(auf der Rechtsachse)  $n$  im Bereich von 10 bis 100 variiert. Wieder wird 1000-mal die binomialverteilte Zufallsgröße  $H$  und damit das zugehörige WALD-Intervall ermittelt. Auch hier irritieren die starken Oszillationen der Überdeckungs-Häufigkeit: Die Erwartung, dass sich diese bei vergrößerndem  $n$  der Sicherheits-Wahrscheinlichkeit  $\delta$  nähern, ist offenbar falsch.

Allerdings ist das WILSON-Intervall hier auch nicht besser, wie Abb. 6 zeigt (und was bei Brown et al. nicht steht). Die Oszillationen rühren von der Diskrettheit der Binomialverteilung her. Abb. 5 ist demnach für sich allein noch kein Argument gegen WALD-Intervalle.

### 3.7 Verbesserungen des WALD-Intervalls

Eine deutliche Verbesserung der Überdeckungs-Häufigkeiten erreicht man beim WALD-Intervall, wenn man in (4) die gemessene Häufigkeit  $H$  durch  $H + \frac{k^2}{2}$  und  $n$  durch  $n + k^2$  ersetzt (Brown et al. 2001;

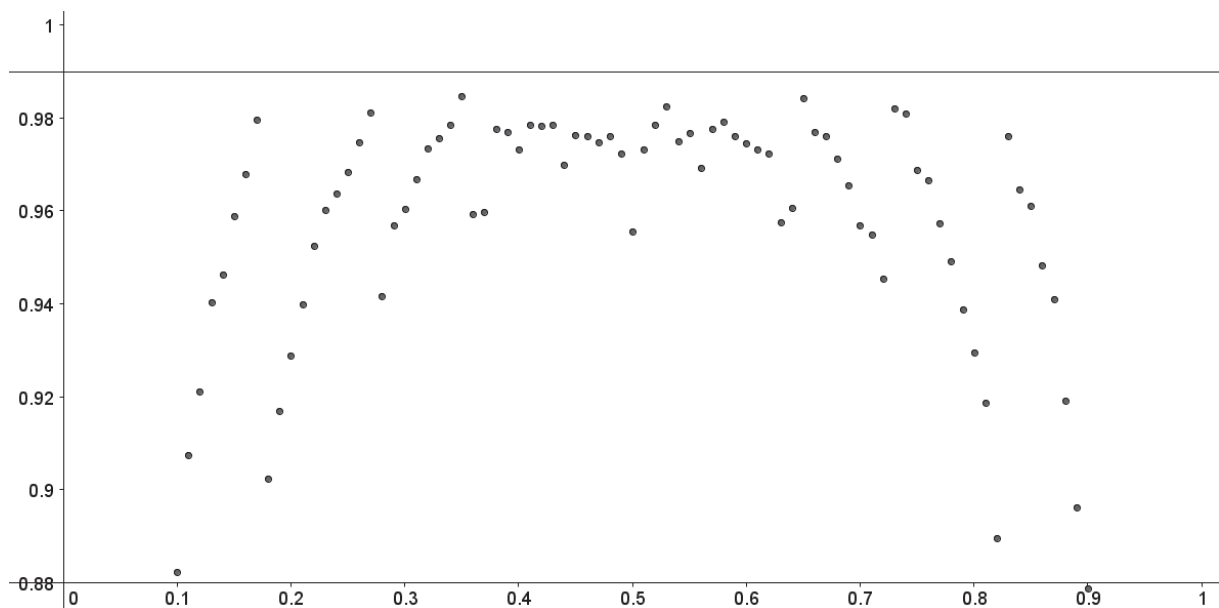


Abb. 4: Überdeckungs-Häufigkeit der WALD-Konfidenz-Intervalle für für  $\delta = 0,99$  und  $n = 20$

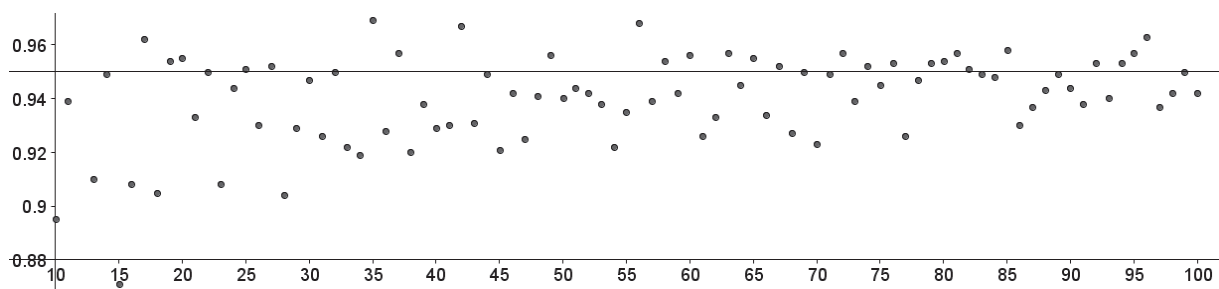


Abb. 5: Überdeckungs-Häufigkeit der WALD-Konfidenz-Intervalle für für  $\delta = 0,95$  und  $p = 0,5$ . Auf der Rechtsachse ist  $n$  abgetragen.



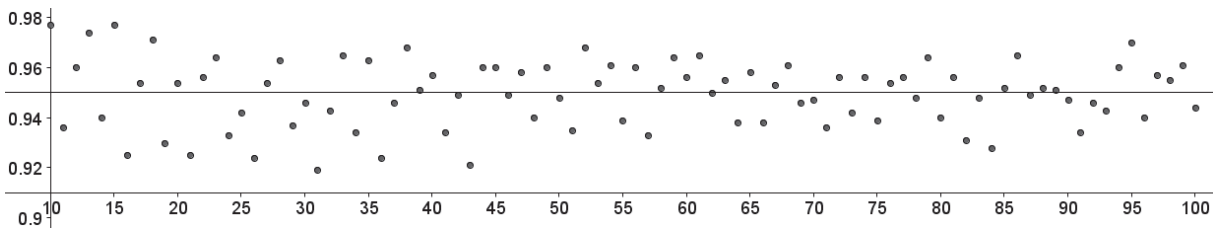


Abb. 6: Überdeckungs-Häufigkeit der WILSON-Konfidenz-Intervalle für  $\delta = 0,95$  und  $p = 0,5$ . Auf der Rechtsachse ist  $n$  abgetragen.

S. 108); für  $k = 2$  heißt das: „Ergänze 2 Erfolge und 2 Misserfolge“. Dies Intervall wird nach Alan Agresti und Brent A. Coull (1998) benannt; es ist nicht mehr symmetrisch zu  $h$ . Allerdings wird man im Schulunterricht kaum derart intensiv auf das Seitenthema „Verbesserungen des Konfidenz-Intervalls“ eingehen wollen, sondern sich auf die grundsätzliche Vorgehensweise beschränken.

Pires und Amado (2008) haben in einer sehr feinen und umfassenden Untersuchung 20 verschiedene Konfidenz-Intervalle auch im Hinblick auf ihre Überdeckungs-Häufigkeiten untersucht und kommen zum Ergebnis (S. 173, 178 und 185), dass eine Stetigkeits-Korrektur zusammen mit einer sinnvollen Behandlung der möglichen Intervall-Länge 0 zu einem modifizierten WALD-Intervall führt, das bezüglich der Überdeckungs-Häufigkeit schon für  $n = 50$  besser abschneidet als das AGRESTI/COULL-Intervall und sogar auch besser als das WILSON-Intervall.

### 3.8 Fazit

Sowohl das in 3.3 erläuterte konzeptuelle Problem, beim Zugang nach WALD zwischen Konfidenz-Intervall und Prognose-Intervall zu unterscheiden, als auch die deutlichen Abweichungen zwischen Überdeckungs-Häufigkeit und Sicherheits-Wahrscheinlichkeit sprechen gegen die Verwendung des WALD-Intervalls im Schulunterricht.

## 4. Das Konfidenz-Intervall nach CLOPPER/PEARSON

### 4.1 Vorgehensweise

Bei den nach (3) und (4) bestimmten Konfidenz-Intervallen spielte bei deren Herleitung die Normalapproximation eine wesentliche Rolle. Rechnet man stattdessen mit der Binomialverteilung, so muss man für  $1 \leq H \leq n - 1$  die beiden Gleichungen

$$\sum_{i=0}^H \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} = \frac{1-\delta}{2}$$

$$= \sum_{i=H}^n \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} \quad (5)$$

jeweils nach  $p$  auflösen. Man kann beweisen, dass dann die Überdeckungs-Wahrscheinlichkeit mindestens so groß wie die Sicherheits-Wahrscheinlichkeit ist (vgl. Henze 2000; Kap. 28.6) – der Beweis ist nicht ganz einfach zu verstehen. Eine entsprechende Aussage wird bei hinreichend häufiger Wiederholung auch für Überdeckungs-Häufigkeiten gelten.

Dies Konfidenz-Intervall ist nach C. J. CLOPPER und EGON S. PEARSON (1934) benannt.

Würde man statt (5) die Gleichungen

$$\sum_{i=0}^{H-1} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} = \frac{1-\delta}{2}$$

$$= \sum_{i=H+1}^n \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

zur Konstruktion des Konfidenz-Intervalls verwenden, so wäre die Überdeckungs-Häufigkeit nur höchstens so groß wie  $\delta$ . Dies kann mit GeoGebra 4.2 nur exemplarisch verifiziert werden.

Die Ermittlung von  $p$  lässt sich etwa mit einem Computer-Algebra-System durchführen; vgl. Meyer (2009). GeoGebra 4.2 bietet dazu nur die (sicherlich nicht schulnahen) Quantile der F-Verteilung an; deren Zusammenhang zu (5) wird in Agresti und Coull (1998; S. 119) angegeben; die Andeutung einer Begründung findet sich etwa bei Leemis/Trivedi (1996; Appendix A). Ist man nur an einzelnen Werten interessiert, lässt sich  $p$  in GeoGebra natürlich mit einem Schieberegler ermitteln.

Die Abb. 7 und 8 illustrieren das Resultat für  $n = 100$  und  $\delta = 0,9$  und  $H = 20$ . Das Konfidenz-Intervall ist [13,666 %; 27,72 %].

Es ist nicht zu  $h$  symmetrisch, was wegen der fehlenden Symmetrie der Binomialverteilung auch nicht zu erwarten gewesen wäre.

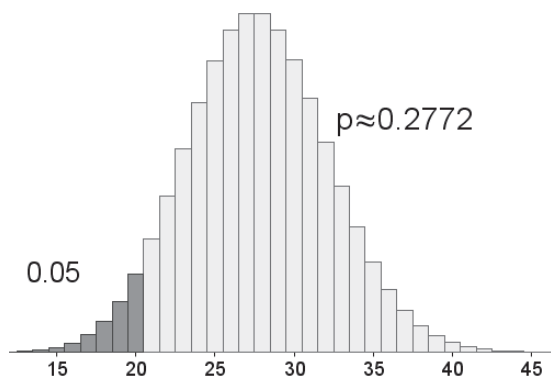


Abb. 7: Zur oberen Grenze des CLOPPER/PEARSON-Intervalls

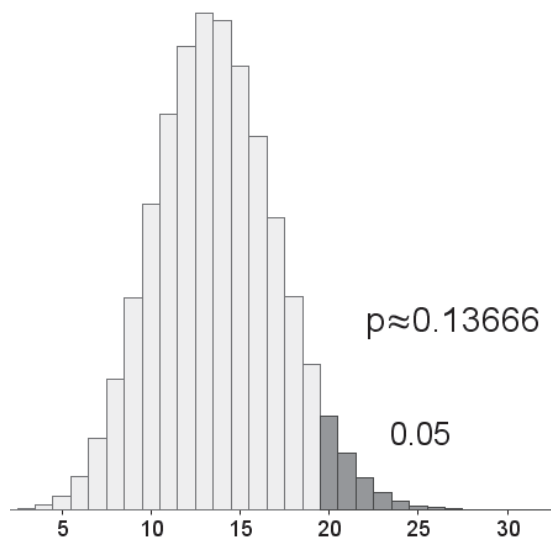


Abb. 8: Zur unteren Grenze des CLOPPER/PEARSON-Intervalls

## 4.2 Zur Überdeckungs-Häufigkeit

Naiverweise könnte man vermuten, dass das aus den Größen  $H$  und  $\delta$  konstruierte CLOPPER/PEARSON-Intervall mit der Wahrscheinlichkeit  $\delta$  den wahren Wert  $p$  überdeckt. Aufgrund der Diskretheit der Binomialverteilung ist diese Vermutung zu naiv. Da GeoGebra 4.2 nicht die Möglichkeit bietet, die erwähnten Quantile der F-Verteilung funktional zu verwenden, wird für die folgende Abbildung auf die Software R (mit dem Paket PropCIs) ausgewichen.

Abb. 9 zeigt die Überdeckungs-Häufigkeiten für  $n = 100$  und  $\delta = 0,9$  und 5.000 Durchgänge. Auf der Rechtsachse sind wieder die wahren Parameter  $p$  aufgetragen, aus denen jeweils 5.000-mal eine binomialverteilte Zufallszahl  $H$  und daraus das CLOPPER/PEARSON-Konfidenz-Intervall ermittelt wurde.

Das Resultat der Abb. 9 ist in Einklang mit Agresti und Coull (1998; S. 119) und Brown et al. (2001; S. 113); diese Autoren führen aus:

Die Überdeckungs-Häufigkeit ist mindestens so groß wie  $\delta$ ; sie kann jedoch deutlich höher sein als  $\delta$  (was man mit der Diskretheit der Binomialverteilung begründen kann), falls  $n$  nicht sehr groß ist. Daher sollte man dies Intervall nicht für die statistische Praxis verwenden, es sei denn, dass  $\delta$  aus inhaltlichen Gründen nicht unterboten werden darf.

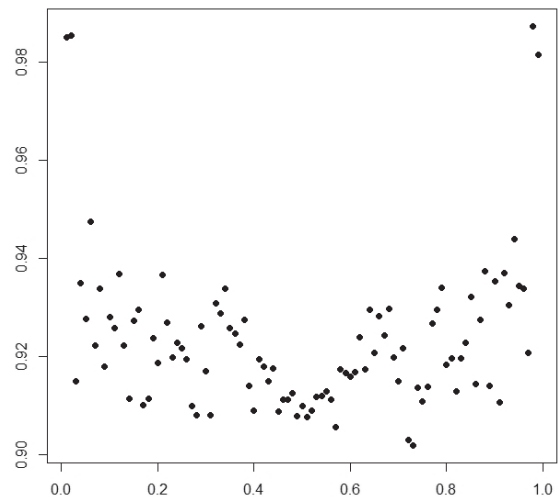


Abb. 9: Überdeckungs-Häufigkeit der CLOPPER/PEARSON-Konfidenz-Intervalle nach (6) für  $\delta = 0,9$  und  $n = 100$

## 4.3 Fazit

Im Schulunterricht wird man kaum die Zeit haben, neben dem WILSON- auch das CLOPPER/PEARSON-Intervall zu behandeln. Dazu kommt als weiterer Hinderungsgrund die (bislang) unzureichende Behandlung mit schulnaher Technologie (zu der R *nicht* gehört!) sowie das Problem, (5) einzusehen. Die gegenüber dem WILSON-Intervall zu große Überdeckungs-Häufigkeit liefert ebenfalls kein Motiv für CLOPPER/PEARSON.

## 5 Die drei Intervalle im Vergleich

In Abb. 10 wurde  $\delta = 0,9$  und  $n = 25$  gewählt. Auf der Rechtsachse sind relative Häufigkeiten  $h$  abgetragen und nach oben die zugehörigen Konfidenz-Intervalle. Links (und etwas nach links verschoben) sieht man jeweils die exakten CLOPPER/PEARSON-Intervalle gemäß der Binomialverteilung, dann das WALD-Intervall gemäß (4) und ganz rechts das WILSON-Intervall nach (3). Abb. 11 zeigt den gleichen Sachverhalt für  $n = 100$ .

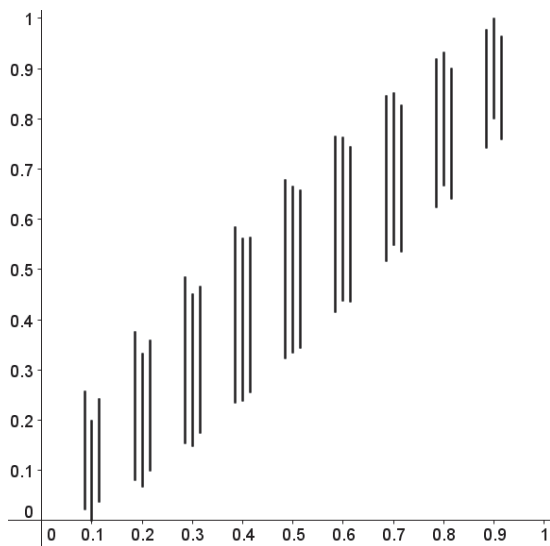


Abb. 10 Verschiedene Konfidenz-Intervalle für  $n = 25$  und  $\delta = 0,9$  zu verschiedenen relativen Häufigkeiten; von links jeweils CLOPPER/PEARSON, WALD, WILSON.

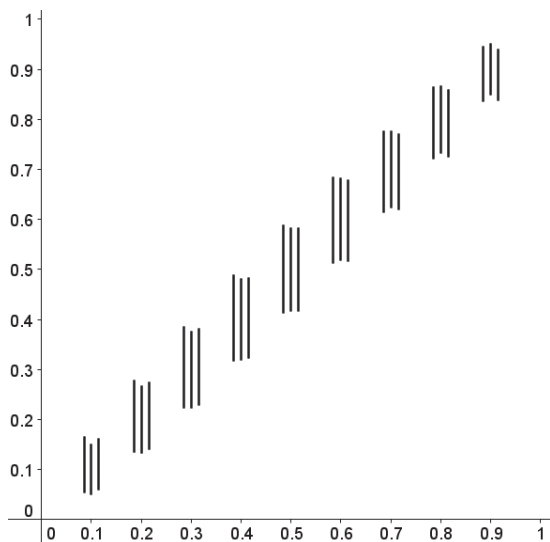


Abb. 11 wie Abb. 10, nun aber für  $n = 100$ .

Es ist deutlich zu sehen, dass sich die drei Konfidenz-Intervalle für größeres  $n$  numerisch nur wenig unterscheiden, so dass die Frage, welches denn nun das „richtige“ sei, an Relevanz verliert. Beide Abbildungen zeigen auch, dass die Breiten der Konfidenz-Intervalle sich nicht dramatisch unterscheiden.

Dagegen unterscheiden sich die drei Konfidenz-Intervalle in ihren Überdeckungs-Häufigkeiten: WILSON schneidet am besten ab, bei CLOPPER/PEARSON ist die Überdeckungs-Häufigkeit größer als die Sicherheits-Wahrscheinlichkeit und oftmals deutlich

größer, und bei WALD ist die Überdeckungs-Häufigkeit häufig deutlich kleiner als die Sicherheits-Wahrscheinlichkeit.

## 6 Schlussbemerkung

Es gibt noch weitere Typen von Konfidenz-Intervallen, die jedoch nicht mehr schulnah sind. Die bequemere Berechenbarkeit des WALD-Intervalls ist angesichts der zur Verfügung stehenden Technologie kein überzeugendes Argument mehr, so dass sich die Empfehlung von Brown et al. (2001) verallgemeinern lässt: Man nehme für den Schulunterricht stets das WILSON-Konfidenz-Intervall.

## Dank

Dem mir unbekanntem Hochschul-Gutachter möchte ich herzlich für seine sehr sorgfältige Kritik an einer früheren Version dieses Artikels danken – und auch dafür, dass er mich ein wesentliches Manko selbst hat entdecken lassen.

## Literatur

- Agresti, A.; Coull, B. A. (1998): Approximate is better than „exact“ for interval estimation of binomial proportions. In: *American Statistician* Vol. 52, S. 119–126.
- Brown, L. D.; Cai, T. T.; DasGupta, A. (2001): Interval estimation for a binomial proportion. In: *Statistical Science* Vol. 16, S. 101–133.
- Engel, A. (1987): Stochastik. Stuttgart: Klett Verlag.
- Henze, N. (2000): Stochastik für Einsteiger: Vieweg.
- Lergenmüller, A.; Schmidt, G; Krüger, K. (Hrsg.) (2012): Neue Wege Stochastik: Schroedel.
- Lemis, L. M.; Trivedi, K. S. (1996): A comparison of approximate interval estimators for the Bernoulli parameter. In: *American Statistician* Vol. 50, S. 63 ff.
- Meyer, J. (2009): Konfidenzintervalle und CAS. In: *Computeralgebra-Rundbrief* Vol. 44, S. 18–20.
- Meyer, J. (2011): Visualisierung stochastischer Inhalte mit GeoGebra. In: *Stochastik in der Schule* Vol. 31, S. 18–22.
- Pires, A. M.; Amado, C. (2008): Interval estimators for a binomial proportion: Comparison of 20 methods. In: *REVSTAT – Statistical Journal* Vol. 6, number 2, 165–197.

## Anschrift des Verfassers

Dr. Jörg Meyer, Hameln  
J.M.Meyer@t-online.de