Schwierigkeiten mit Konfidenzintervallen? In der Tat!

RAPHAEL DIEPGEN, BOCHUM

Zusammenfassung: *Es wird kritisch zu dem SiS-Beitrag „Schwierigkeiten mit Konfidenzintervallen" von Jörg MEYER (2013) Stellung genommen. Insbesondere wird seiner Empfehlung widersprochen, im Schulunterricht nicht Konfidenzintervalle nach WALD zu behandeln.*

1 Einleitung

In seinem informativen Beitrag im letzten SiS-Heft von 2013 diskutiert Jörg MEYER aus didaktischer Perspektive im Wesentlichen Konfidenzintervalle nach WILSON im Vergleich zu Konfidenzintervallen nach WALD, und zwar Konfidenzintervalle jeweils für eine unbekannte Wahrscheinlichkeit p, basierend auf der Häufigkeit H in einer Zufallsstichprobe vom Umfang n. Deren Binomialverteilung mit den Parametern n und p wird in beiden Fällen durch die Normalverteilung approximiert.

2 Konzept des Konfidenzintervall

Überschreitet die (angenähert) standardnormalverteilte Zufallsvariable $\frac{H-np}{\sqrt{np(1-p)}}$ betragsmäßig den kritischen Wert k mit der vorgewählten Wahrscheinlichkeit δ, ist also in statistiküblicher Notation $k=z_{\delta/2}$, ergibt sich für die relative Häufigkeit $h=H/n$ die – von p selbstverständlich abhängige – Wahrscheinlichkeit $P_p\left(|h-p|\le k\sqrt{\frac{p(1-p}{n}}\right)=1-\delta$, im Intervall mit den Grenzen $p\pm k\sqrt{\frac{p(1-p)}{n}}$ zu liegen, gemeinhin 1–δ-Prognoseintervall genannt. Fragt man nun für eine beobachtete relative Häufigkeit h nach der Menge der hypothetischen Parameterwerte p, für die h in dem zu p gehörenden 1–δ-Prognoseintervall läge, wäre diese offensichtlich durch ein Intervall gegeben mit Grenzen definiert durch $|h-p|=k\sqrt{\frac{p(1-p)}{n}}$. Da für jedes hypothetische p aus diesem Intervall h mit Wahrscheinlichkeit 1–δ höchstens um $k\sqrt{\frac{p(1-p)}{n}}$ von p entfernt liegt, liegt umgekehrt – für jedes denkbare p – p mit Wahrscheinlichkeit 1–δ höchstens um $k\sqrt{\frac{p(1-p)}{n}}$ von h entfernt. Das fragliche – vom zufälligen h abhängige – Intervall überdeckt also mit der Wahrscheinlichkeit 1–δ das unbekannte p. Soweit die grundlegende Idee des Konfidenzintervalles.

Bestimmt man nun die Grenzen dieses Intervalls durch Lösen obiger – letztlich quadratischen – Gleichung, so ergeben sich die – um h nicht symmetrischen – Grenzen $\frac{h+\frac{k^2}{2n}\pm\frac{k}{\sqrt{n}}\sqrt{h(1-h)+\frac{k^2}{4n}}}{1+\frac{k^2}{n}}$; das ist das Konfidenzintervall nach WILSON. Ersetzt man in diesen Grenztermen – bei großem n – jeweils k^2/n durch 0, erhält man das Konfidenzintervall nach WALD mit den – um h nun symmetrischen – Grenzen $h\pm k\sqrt{\frac{h(1-h)}{n}}$. Dieses Konfidenzintervall ergibt sich äquivalent natürlich auch, wenn man in der Ausgangsbedingung $|h-p|=k\sqrt{\frac{p(1-p)}{n}}$ auf der rechten Seite jeweils p durch h ersetzt. Soweit zur Begrifflichkeit.

3 Didaktische Bewertung

MEYER plädiert nun dafür, im Unterricht das Konfidenzintervall nach WILSON zu verwenden. Die Schwierigkeiten der dafür erforderlichen Lösung der quadratischen Gleichung „zu Fuß" könne man „einfach" durch Einsatz eines graphischen Taschenrechners umgehen. Gegen die Behandlung des WALD-Konfidenzintervalles im Unterricht sprächen zwei Argumente:

1. Deutliche Abweichungen zwischen nomineller Sicherheitswahrscheinlichkeit 1–δ und tatsächlicher Überdeckungshäufigkeit beim WALD-Konfidenzintervall.
2. Das strukturelle oder konzeptuelle Problem der bloß formal gleichen Vorgehensweise bei der Bestimmung von Prognoseintervall einerseits, von WALD-Konfidenzintervall andererseits.

3.1 Zu geringe Überdeckungshäufigkeiten der WALD-Konfidenzintervalle

Hier präsentiert MEYER in seinen Abbildungen 3 bzw. 4 die Ergebnisse jeweils 1000-facher Simulationen der $B(25,p)$- bzw. $B(20,p)$-verteilten Häufigkeit

H zu verschiedenen Werten von p: Dargestellt wird jeweils die relative Häufigkeit (unter den 1000 Simulationen), mit der das in Abhängigkeit vom simulierten H kalkulierte 95 %- bzw. 99 %-WALD-Konfidenzintervall p überdeckt. Hier fällt dann auf, dass diese Überdeckungshäufigkeiten für einen breiten Mittelbereich von p ungefähr der Sicherheitswahrscheinlichkeit 95 % bzw. 99 % (knapp) entsprechen, für extreme Werte von p nahe 0 bzw. 1 aber deutlich darunter abfallen – ein unerwünschtes Phänomen, das entsprechende Simulationen für die WILSON-Konfidenzintervalle nicht zeigen.

Verblüffend ist, worauf MEYER dieses Phänomen – leider ohne nähere Erklärung – attribuiert: „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 zugehörige Binomialverteilung sehr asymmetrisch ist.“ (S. 14) Wie aber könnte der skizzierte Unterschied in den Überdeckungshäufigkeiten zwischen WILSON- und WALD-Konfidenzintervallen jemals mit der – bei WILSON gleichsam berücksichtigten, bei WALD gleichsam ignorierten – Asymmetrie der Binomialverteilung zu tun haben, wo doch beide Konfidenzintervalle in gleicher Weise die asymmetrische Binomialverteilung von Vornherein durch die symmetrische Normalverteilung ersetzen? Das kann doch gar nicht sein. „Kompensiert“ etwa das asymmetrische WILSON-Intervall die durch die Normalverteilungsapproximation erzwungene Symmetrisierung der Verteilung? Hat die Asymmetrie der WILSON-Konfidenzintervalle (für $h\neq 0{,}5$) ihren Grund in der Asymmetrie der Binomialverteilung (für $p\neq 0{,}5$) – der dann gleichsam von den WALD-Konfidenzintervallen bei Strafe zumindest am Rande zu kleiner Überdeckungshäufigkeiten missachtet würde? Nein. Die – exakten – WILSON-Konfidenzintervalle sind nicht asymmetrisch wegen der Asymmetrie der Binomialverteilung, von der sie der ersetzenden Normalverteilung wegen gar nichts „wissen“ können, sondern sie sind asymmetrisch, weil die Varianz $p(1-p)$ der zugrundeliegenden Bernoulli-Verteilung eine (quadratische) Funktion ihres Mittelwertes p ist. Dies – und nicht irgendeine Asymmetrie der Binomialverteilung – führt dazu, dass für ein gegebenes $h<0{,}5$ das weitest entfernte p mit $|h-p|\leq k\sqrt{\frac{p(1-p)}{n}}$ zur Linken weniger weit entfernt ist als zur Rechten (und für ein $h>0{,}5$ vice versa), so dass das resultierende WILSON-Konfidenzintervall nicht symmetrisch um h liegt. Für $h=0{,}1$ (und irgendein k und n) illustriert diese Asymmetrie die Abb. 1[1] der drei Funktionen (1) $f(p)=k\cdot\sqrt{\frac{p(1-p)}{n}}$, (2) $f(p)=|h-p|= h-p$ für $p\leq h$ und (3) $f(p) = |h-p| = -h+p$ für $p\geq h$:

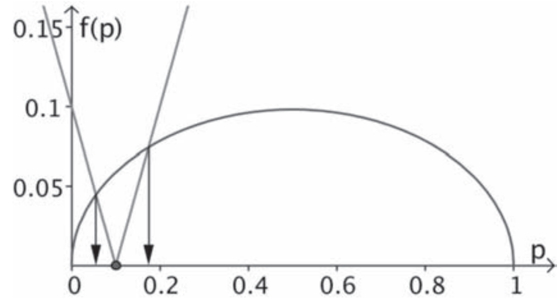

Abb. 1

Die Asymmetrie der WILSON-Konfidenzintervalle hat ihren Grund nicht in einer Asymmetrie der Binomialverteilung, die von der zwischengeschalteten Normalverteilung ohnehin nicht erhalten wird, sondern in einer bestimmten funktionellen Abhängigkeit der die Breite von Prognoseintervallen determinierenden Varianz der Binomialverteilung (und der approximativ passenden Normalverteilung) vom zu schätzenden Parameter p. Dies ist nun aber eine besondere und wenig verallgemeinerungsfähige Eigenschaft der Binomialverteilung. In den Fällen – und diese sind wohl viel typischer –, in denen die Varianz des Schätzers unabhängig vom zu schätzenden Parameter ist – prominentes Beispiel: Konfidenzintervall für den Erwartungswert μ einer Variablen mit davon unabhängiger Varianz σ^2 – gibt es diese Problematik nicht: Die Konfidenzintervalle sind dort grundsätzlich symmetrisch. Die ausdrückliche Ermahnung von MEYER (S. 14), man solle ein Konfidenzintervall nicht als „Schätzwert±Fehler“ auffassen, dürfte daher nur selten treffen – und so für die Schule eher entbehrlich sein.

Wenn also der von MEYER monierte gravierende Abfall der Überdeckungshäufigkeiten von WALD-Konfidenzintervallen für extremes p nahe 0 oder 1 seinen Grund wohl kaum, wie vom ihm angedeutet, in der Symmetrie dieser Intervalle im Kontrast zur Asymmetrie der entsprechenden Binomialverteilungen haben dürfte, fragt sich nach einer anderen Erklärung für diesen Befund. Nun, schauen wir etwa auf die von MEYER in seiner Abbildung 3 für $p\approx 0{,}01$ % berichtete Überdeckungshäufigkeit von nur etwa 22 %, bezogen auf die Simulation von 95 %-WALD-Konfidenzintervallen zur (mäßigen) Stichprobengröße $n=25$. Die Wahrscheinlichkeit, dass die $B(25;\, 0{,}01)$-verteilte Häufigkeit H den Wert 0 annimmt und damit das WALD-Konfidenzintervall auf [0;0] mit der Breite 0 schrumpft und – außer der

0 selbst – gar nichts überdeckt, insbesondere also auch nicht p, beträgt mit etwa $0{,}99^{25}$ rund 78 %, also ziemlich genau die Gegenwahrscheinlichkeit zur angegebenen Überdeckungshäufigkeit. Spielt man diese Kalkulation für die von MEYER in seinen Abbildungen 3 und 4 präsentierten, in der Nähe von 0 und 1 stark abfallenden Überdeckungshäufigkeiten für die anderen simulierten extremen Werte von p durch, findet man ähnliche Übereinstimmungen. Kurzum: Der von MEYER monierte und auf die Asymmetrie der Binomialverteilung attribuierte Abfall der Überdeckungshäufigkeiten der symmetrischen WALD-Intervalle für sich 0 oder 1 näherndes p dürfte seinen wesentlichen Grund in der gleichzeitigen Zunahme der Wahrscheinlichkeiten von (außer der 0 bzw. der 1) gar nichts überdeckenden WALD-Konfidenzintervallen der Breite 0 haben.

Könnte dies das MEYERsche Verdikt gegen die WALD-Konfidenzintervalle rechtfertigen? Wohl kaum. Wie könnte man jemals ernsthaft von einem offensichtlich abwegigen „uneigentlichen" Konfidenzintervall [0;0] bzw. [1;1] in die Irre geführt werden? Wer würde jemals glauben, ein solches Intervall überdecke den unbekannten Parameter, der das statistische Geschehen ja zu einem deterministischen machen würde? Übrigens: Die bedingten Überdeckungswahrscheinlichkeiten beim WALD-Konfidenzintervall unter der Bedingung, dass es nicht uneigentlich ist, dürften der nominellen Sicherheitswahrscheinlichkeit durchaus genügen. Am Beispiel von oben (p=0,01, n=25, 1–δ=0,95, also k=1,96): Hier überdeckt das WALD-Konfidenzintervall p=0,01 genau dann, wenn H=1, H=2 oder H=3 ist, was eine Wahrscheinlichkeit von 0,2207 hat; H=0 und damit das uneigentliche Konfidenzintervall [0;0] hat die Wahrscheinlichkeit 0,7778. Also beträgt die Wahrscheinlichkeit, dass das WALD-Konfidenzintervall p=0,01 überdeckt unter der Bedingung, dass es nicht das – als nichtssagend sofort erkennbare – uneigentliche Intervall [0;0] ist, beachtliche 0,9934.

Dass WALD-Konfidenzintervalle für Parameter p nahe bei 0 oder 1 häufig uneigentlich werden, ist halt der Preis für den schätzenden Ersatz von p durch h auf der rechten Seite der Ausgangsbedingung. Die wegen dieser häufigen uneigentlichen Intervalle geringe Überdeckungshäufigkeit im Randbereich von p ist aber – ganz anders als eine mangelnde Überdeckungshäufigkeit trotz überwiegend eigentlicher Intervalle mit positiver Breite im mittleren Bereich von p – wenig störend, weil die uneigentlichen Intervalle [0;0] und [1;1] in Unwissenheit zurücklassen, nicht aber in die Irre führen. Die „Irrtumswahrscheinlichkeit" dürfte beim WALD-Konfidenzintervall – auch für extremes p – kaum größer sein als beim WILSON-Konfidenzintervall.[2]

Fazit: Das vermeintliche – für großes n ohnehin verschwindende – „Problem" der im Randbereich zu geringen Überdeckungshäufigkeiten der WALD-Konfidenzintervalle entpuppt sich als kaum ernsthaft störende Folge der Möglichkeit der uneigentlichen Konfidenzintervalle [0;0] bzw. [1;1]. Dies disqualifiziert WALD-Konfidenzintervalle für den Unterricht sicherlich *nicht*.

Gibt es auf der anderen Seite – außer seiner rechnerischen Einfachheit (und der Befreiung von der Last der komplizierten oder technologieabhängigen Lösung einer quadratischen Gleichung) im Vergleich zum WILSON-Konfidenzintervall – didaktische Vorteile des WALD-Konfidenzintervalles? Meines Erachtens: Ja. Diese werden allerdings kaum sichtbar, solange man – wie MEYER – das WALD-Konfidenzintervall über das „vereinfachende" Nullsetzen von k^2/n in der WILSON-Formel respektive den Ersatz der unbekannten Wahrscheinlichkeit p durch die beobachtete relative Häufigkeit h rechts in der Ausgangsbedingung $|h-p| = k\sqrt{\dfrac{p(1-p)}{n}}$ einführt.

Der eigentliche Gag des WALD-Konfidenzintervalles ist doch aber gerade der, dass man die – („nur") bei der Binomialverteilung störender Weise von p abhängige – Varianz auf der rechten Seite der Ausgangsbedingung aus den Daten schätzt. Man entledigt sich so einer spezifischen – und damit wenig allgemeinbildenden – Schwierigkeit bei der Binomialverteilung (und bezahlt dafür gerne durch die wenig störende Möglichkeit irrelevanter uneigentlicher Konfidenzintervalle und halt die bei jeder Schätzung üblichen Ungenauigkeiten und Unvollkommenheiten). Das ist aber für die Statistik ein typisches Vorgehen – typischer vermutlich als das ohnehin nur selten mögliche exakte Rechnen à la WILSON-Konfidenzintervall: Sucht man ein Konfidenzintervall für den unbekannten Erwartungswert μ einer Variablen mit davon unabhängiger unbekannter Varianz σ^2, bleibt einem gar nichts anderes übrig, als die Varianz à la WALD aus den Daten zu schätzen; ein Vorgehen à la WILSON ist hier schlicht nicht möglich. Damit ist aber grundsätzlich das WALD-Konfidenzintervall mit seiner integrierten Varianz*schätzung* das allgemeinere Konzept – und daher in der Schule wohl vorzuziehen. Es macht wenig Sinn, sich beim Unterricht zum Konfidenzintervall auf Besonderheiten zu konzentrieren, die überhaupt nur für die Binomialverteilung gelten. Es geht im Schulunterricht allenfalls um das Konzept des Konfidenzintervalles im Allgemeinen, nicht um spezielle Konfidenzintervalle mit maximaler Übereinstimmung zwischen Überdeckungshäufigkeit und nomineller Sicherheitswahrscheinlichkeit bei einer bestimmten Verteilung. Die Binomialverteilung ist in diesem Unterricht nur ein Beispiel – und damit interessant nur im Hinblick auf ihre verallgemeinerungsfähigen Aspekte: Gerade diese werden aber eher vom WALD-Konfidenzintervall repräsentiert als vom WILSON-Konfidenzintervall.

3.2 Das strukturelle Problem der bloß formal gleichen Vorgehensweise bei Prognoseintervallen und WALD-Konfidenzintervallen

Als zweiten Grund für seine Ablehnung der WALD-Konfidenzintervalle nennt MEYER die – bei Schülern Verwirrung stiftende – bloß formal gleiche Vorgehensweise zu Prognoseintervallen. Hie die Formel $h \pm k\sqrt{\frac{h(1-h)}{n}}$, da $p \pm k\sqrt{\frac{p(1-p)}{n}}$. Nun fragt sich, worin die Verwirrung eigentlich bestehen soll: Hie sucht man jeweils zu einer beobachteten relativen Häufigkeit h einen Bereich mutmaßlicher Wahrscheinlichkeitsparameter p; dort sucht man zu einem gegebenen Wahrscheinlichkeitsparameter p einen Bereich wahrscheinlicher relativer Häufigkeiten h. Wie könnte ein Schüler diese beiden völlig verschiedenen Fragestellungen oder Situationen – allgemein: Suche nach einem Intervall mutmaßlicher Parameter verträglich mit dem beobachteten Wert eines Schätzers versus Suche nach einem Intervall wahrscheinlicher Werte einer Zufallsvariablen bei bekanntem Parameter – jemals verwechseln nur dadurch, dass die Lösungsverfahren formale Parallelen haben? Vor allem: Wie könnte er jemals, „im echten Leben“ in der einen Situation befindlich irrtümlich wähnen, in der anderen Situation zu sein? Droht da wirklich ein ernsthaftes Problem? (Oder ist dies nur ein „Problem“ in der lebensfernen Kunstwelt formelhaft zu bearbeitender, nicht problemorientierter Mathematikaufgaben?) Und sollte man diesem Problem mit einer didaktischen Maßnahme begegnen, die ohnehin nur im speziellen Fall der Binomialverteilung greift? Nach dem oben in 3.1 Gesagten dürfte nun ohne ausführliche Erläuterungen klar sein: Ist – anders als im Spezialfall der Binomialverteilung – die Varianz $\sigma^2_{\hat{\pi}}$ eines (approximativ normalverteilten) Schätzers $\hat{\pi}$ unabhängig vom zu schätzenden Parameter π, ergibt sich eine *grundsätzlich unauflösbare* formale Parallelität zwischen Konfidenzintervall und Prognoseintervall aufgrund der für beide einheitlichen Ausgangsgleichung $|\hat{\pi} - \pi| = k\sigma_{\hat{\pi}}$, die hie nach π, dort nach $\hat{\pi}$ aufgelöst wird. Diese grundsätzlich unauflösbare Parallelität ist eigentlich auch nicht nur eine bloß „formale“, denn sie gründet in der grundsätzlichen begrifflichen Verschränkung von Konfidenzintervallen mit Prognoseintervallen, die schließlich im „Normalfall“ der vom fraglichen Parameter unabhängigen einheitlichen Schätzervarianz beide gleiche Breiten haben, und zwar zur Linken genauso lang wie zur Rechten. Auch hier kann die Ablehnung der WALD-Konfidenzintervalle für den Schulunterricht kaum überzeugen.

4 Weitere didaktische Überlegungen

„‚Die gesuchte Einzelerfolgs-Wahrscheinlichkeit liegt mit Wahrscheinlichkeit 95 %‘“ im 95 %-Konfidenzintervall. Es ist wohl nicht untypisch, wie „technisch“ und „oberflächlich“ auch MEYER (S. 11) mit dieser für jeden frequentistisch orientierten Unterricht zum Konfidenzintervall ja so typischen „Schüler-Schwierigkeit“ umgeht: Man solle auf den nicht zufälligen Charakter von p hinweisen und in entsprechenden Diagrammen visualisieren, dass von vielen zufallsabhängigen Realisationen eines Konfidenzintervalles eben ein Anteil von etwa δ den festen wahren Parameter nicht überdeckt. Nur: Die von MEYER wie üblich nur formal erledigte „Schüler-Schwierigkeit“ deutet auf eine echte konzeptuelle Schwierigkeit des frequentistischen Konfidenzintervalles hin, wie sie von Seiten der Bayesianischen Statistik schon immer kritisiert wird. Es ist ja nun tatsächlich umstritten, ob das Konfidenzintervall ein rationales Verfahren ist – trotz seiner weiten Verbreitung. Diese Grundlagenproblematik sollte man dann aber auch im allgemeinbildenden und aufklärenden Unterricht ausführlich thematisieren: Worin soll denn der Sinn eines 95 %-Konfidenzintervalles bestehen, wenn man nach seiner Realisierung eben nicht sagen kann, der wahre Parameter p liege mit Wahrscheinlichkeit 95 % in dem realisierten Intervall $[a,b]$ – wie es die Schüler (und vielleicht alle vernünftigen Menschen) doch offensichtlich gerne täten? (Warum und was) Soll oder kann man dann noch darauf wetten, dass p in $[a,b]$ liegt? Die ernsthafte Diskussion solcher Fragen erfordert dann auch sicherlich, dass man im Unterricht die Bayessche Alternative zum frequentistischen Konfidenzintervall thematisiert, also die Modifikation einer – zur Vereinfachung ins Diskrete vergröberten – a priori-Verteilung von p zur a posteriori-Verteilung nach Beobachtung einer relativen Häufigkeit h. Die Thematisierung solcher grundsätzlicher Fragen erscheint mir allemal wichtiger als die mathematisch-technischen Feinheiten à la WILSON versus WALD bei Konfidenzintervallen nur zur Binomialverteilung.

Wer den Schülern die skizzierte „Schüler-Schwierigkeit“ beim Konfidenzintervall nur formal „austreibt“, der nimmt ihnen allenfalls ihren Status als Laien im Sinne der schönen Definition: „Laien sind die Menschen, die noch nicht gelernt haben, nur die Fragen zu stellen, auf die die (hier: frequentistisch orientierten) Experten eine Antwort haben.“ Und letztlich bleibt dann immer noch die Frage, wie nachhaltig das formale Austreiben der „Schüler-Schwierigkeit“ bleibt, ob sich nicht alsbald die „Sinn gebende“ und „natürliche“ nichtfrequentistische Missinterpretation der Sicherheitswahrscheinlichkeit als Bayessche Wahrscheinlichkeit über die Lage eines Parameters wieder einstellt. Es sei an die nicht ganz unplausible These erinnert, die Verfahren frequentistischer Statistik hätten sich in vielen Bereichen wohl nur deshalb als häufig ritualisierter Standard etabliert, weil sie von den Anwendern im bayesschen Sinne miss- oder besser: überinterpretiert werden.

Um Zeit zu schaffen, im Unterricht zum Konfidenzintervall solche grundlegenden Fragen zu reflektieren, sollte man den vom Grundsätzlichen ablenkenden mathematisch-technischen Aufwand so gering wie möglich halten. Dazu einige Anregungen:

- Man könnte es bei der abstrakten Definition belassen, jeweils für eine beobachtete relative Häufigkeit h das Intervall der Werte für den Wahrscheinlichkeitsparameter p zu bilden, für die das beobachtete h im $1–\delta$-Prognoseintervall zu p läge. Dass man die durchprobierende Zusammenstellung dieser Werte dann einem geeigneten Computerprogramm überlassen kann, sofern dieses die $1–\delta$-Prognoseintervalle – exakt oder angenähert – berechnen kann, dürfte auch den Schülern klar sein. Zu verstehen wäre hier lediglich, dass dieses Verfahren mit Wahrscheinlichkeit $1–\delta$ ein Intervall liefert, das das unbekannte p überdeckt – sicherlich für manchen Schüler eine gedanklich-logische Herausforderung. Da in diesem Unterricht keinerlei Formeln zur Berechnung von Konfidenzintervallen entwickelt werden, lassen sich hier grundlegende Erkenntnisse zu Zusammenhängen zwischen dem Niveau $1–\delta$, dem Stichprobenumfang n und der Breite des Konfidenzintervalles, also etwa das beliebte Eins-durch-Wurzel-n-Gesetz, natürlich nicht gewinnen.
- Man könnte zur Vereinfachung in der Ausgangsbedingung $|h-p| \le k\sqrt{\frac{p(1-p)}{n}}$ den Varianzterm $p(1–p)$ durch seinen Höchstwert 0,25 abschätzen. Das Niveau $1–\delta$ ist dann nur noch das untere Limit für die Sicherheitswahrscheinlichkeit und das Konfidenzintervall wird breiter. Wie gesagt: Es geht im Unterricht um die Grundidee des Konfidenzintervalles – und nicht darum, dass es möglichst schmal ist.
- Man könnte die Approximation durch die Normalverteilung umgehen – also letztlich für den Schüler nicht gänzlich nachvollziehbare Mathematik –, indem man Konfidenzintervalle auf Basis der einfachen Ungleichung von Tschebyscheff konstruiert, nämlich $P_p\left(|h-p| \ge k\sqrt{\frac{p(1-p)}{n}}\right) \le \frac{1}{k^2}$. Man mündet dann – auf einfachem, aber mathematisch für den Schüler vollständig nachvollziehbarem Weg – zu derselben Ausgangsgleichung $|h-p| = k\sqrt{\frac{p(1-p)}{n}}$ wie oben, nur diesmal mit dem kritischen Wert $k = 1/\sqrt{\delta}$, der deutlich größer ist als der kritische Wert beim Weg über die Normalverteilung – mit der (im Unterricht tragbaren) Folge deutlich breiterer Konfidenzintervalle. Auch hier ist das Niveau $1–\delta$ nur noch untere Grenze für die Sicherheitswahrscheinlichkeit; auch hier könnte man zusätzlich $p(1–p)$ durch ¼ abschätzen. Nebenbei: Die Konstruktion über die Ungleichung von Tschebyscheff ist an keine Verteilung gebunden, ist also auch dann gangbar, wenn man es nicht mit der Binomialverteilung zu tun hat, auch dann, wenn man außer der – ggf. nur geschätzten – Varianz nichts über die zugrundeliegende Verteilung weiß. Das oben mehrfach erwähnte Konfidenzintervall für den Erwartungswert einer beliebig verteilten Zufallsvariablen ließe sich so im Unterricht einfach ableiten – ohne Berufung auf den Zentralen Grenzwertsatz.

Last not least: Man sollte im Unterricht konsequent mit dem missverständnisträchtigen Begriff „Konfidenzintervall“ nur das Verfahren benennen, nicht ein konkret realisiertes Intervall. (Anders also, als es auch in diesem Beitrag hier geschehen ist. Aber für den Fachmann ergibt sich – anders als für den Schüler – sofort aus dem Kontext, was gemeint ist.) Dass mit diesem Begriff zumeist sowohl ein zufälliges Intervall, als auch eine konkrete Realisation desselben bezeichnet wird, ist sicherlich eine wesentliche Ursache für manche Missverständnisse.

Nachtrag 1

Dass die Überdeckungswahrscheinlichkeiten der WALD-Konfidenzintervalle für p nahe 0 oder 1 geringer sein dürften als die der exakten WILSON-Konfidenzintervalle, zeigen auch folgende Betrachtungen, ausgehend von formalen Beziehungen, die sich mehr oder minder leicht ableiten lassen:

1. Die Breite des WILSON-Konfidenzintervalles ist genau dann größer als die des WALD-Konfidenzintervalles, wenn $\left|h-\frac{1}{2}\right| > \frac{1}{2}\cdot\sqrt{1-\frac{1}{2+k^2/n}}$ ist, wenn also h nicht in der Mitte, sondern in der Nähe von 0 oder 1 liegt – und damit vermutlich auch p. Je breiter aber ein Konfidenzintervall ist, desto größer ist tendenziell seine Trefferwahrscheinlichkeit.
2. Das WALD-Konfidenzintervall liegt nur dann in $[0;1]$, wenn $h \in [1-a;\, a]$ ist mit $a := \frac{1}{1+k^2/n} < 1$. Für extremes h jenseits dieser Limits – und damit tendenziell für extremes p – „verschenkt" also das WALD-Konfidenzintervall Treffermöglichkeiten, da es auch unmögliche Werte kleiner 0 oder größer 1 umfasst.
3. Das WALD-Konfidenzintervall mit Mitte h ist gegenüber dem exakten WILSON-Konfidenzintervall mit Mitte $p_m = a\cdot h + (1-a)\cdot\frac{1}{2}$ umso stärker zu den Rändern hin verschoben, je extremer h – und damit tendenziell p – ist. Denn $\frac{1}{2}-h=\frac{1}{a}\left(\frac{1}{2}-p_m\right)$ mit $\frac{1}{a}>1$.

   Diese Verschiebung ist aber wohl mit einem Verlust an Trefferwahrscheinlichkeit verbunden.

Nachtrag 2

Mein Plädoyer für WALD wurzelt in einem – sachfremden, gleichwohl wirksamen – Gefühl einer doppelten Sympathie:

1. Viel mehr noch als für den Unterricht zum Konfidenzintervall verspricht WALD für den Unterricht über das Hypothesentesten. Würde man in das statistische Hypothesentesten nach Neyman und Pearson in der Version des sequentiellen Testens von WALD einführen (DIEPGEN, 1987), würde man sich elegant der meisten Probleme entledigen, an denen dieser Unterricht seit Jahrzehnten krankt – und die auch wieder einmal im ersten SiS-Heft 2014 lang und breit beklagt wurden. Fast wäre man versucht, den Statistikdidaktikern zuzurufen: „Ihr seht vor lauter Bäumen den WALD nicht mehr!"
2. Als nebenberuflich tätiger Pilot und Luftfahrtpsychologe berühren mich Opfer von Pilotenfehlern irgendwie besonders. Abraham WALD bezahlte einen solchen Fehler mit seinem Leben.

Anmerkungen

1 Ich verdanke Sie einem Gutachter. Danke.

2 Im Nachtrag finden sich auf Anregung eines Gutachters weitere Überlegungen, warum die Konfidenzintervalle nach WALD für extreme p geringere Überdeckungswahrscheinlichkeiten haben dürften als die nach WILSON.

Literatur

Diepgen, R. (1987): Sequentielles Testen – auch didaktisch vielleicht eine gute Alternative. In: *Stochastik in der Schule 7 (2)*, S. 9–25.

Meyer, J. (2013): Schwierigkeiten mit Konfidenzintervallen. In: *Stochastik in der Schule 33 (3)*, S. 10–17.

Anschrift des Verfassers

Raphael Diepgen
Fakultät für Psychologie
Ruhr-Universität Bochum
Universitätsstr. 150
44780 Bochum
raphael.diepgen@rub.de