Experimente zur Einführung von Ideen und Denkweisen statistischer Inferenz im Gymnasium

Manfred Borovcnik, Klagenfurt; Peter Fejes-Tóth, Budapest; Zsuzsanna Jánvári, Budapest; Ödön Vancsó, Budapest

Zusammenfassung: *Das ungarische Gymnasium bereitet auf den Hochschulzugang vor. Die Ausbildung in Stochastik ist auf die beschreibende Statistik beschränkt. Eines der Ziele einer Forschungsgruppe an der Ungarischen Akademie der Wissenschaften ist die Vorbereitung der Reform des Curriculums in Wahrscheinlichkeitsrechnung und Statistik am Gymnasium (Klassenstufen 10–12). In diesem Artikel präsentieren wir Experimente, die Lernende in Gruppenarbeit durchführen können. Durch diese interaktiven Experimente können neue Konzepte zum Wahrscheinlichkeitsbegriff und zur statistischen Denkweise auf eine Art eingeführt werden, die zu unserer Ansicht von den dahinterstehenden Ideen passt; die Vorgangsweise kann als empirisch eingestuft werden. Wir bemühen uns auch, klassische und Bayesianische Sichtweisen zur beurteilenden Statistik schon im Anfangsunterricht einzubringen.*

Schlagwörter: Statistische Inferenz, Wahrscheinlichkeitsverteilungen, Mathematik der Höheren Schule, Interdisziplinärer Ansatz.

1 Einleitung und Hintergrund

Der Aufsatz beschreibt das Design und den Zweck von vier Experimenten. Das erste führt zur Hypothesenprüfung; die daran anschließenden Experimente 2 und 3 schließen den klassischen und den Bayesianischen Ansatz zur beurteilenden Statistik mit ein. Das vierte führt die Normalverteilung und ihre grundlegenden Eigenschaften ein und verbindet beurteilende Statistik mit der Art, wie man in empirischer Forschung Fragen analysiert und Ergebnisse bewertet. In der ungarischen Schulbildung werden traditionell Spiele und Experimente genützt, um mathematische Konzepte und Begriffe aufzubauen; die Kernidee der Begriffe soll sich aus dem Kontext der Spiele ergeben.

Beurteilende Statistik ist kein Teil des Lehrplans der Sekundarschule in Ungarn. Es bestehen aktuell konkrete Pläne zur Reform des nationalen Lehrplans, die eine Einführung grundlegender Begriffe und Methoden der beurteilenden Statistik miteinschließen. Eine Arbeitsgruppe der Ungarischen Akademie der Wissenschaften, der die Autoren angehören, untersucht Möglichkeiten zur Einführung und gestaltet Aufgaben und praktikable Lernwege.

Eines der Haupthindernisse bei der Einführung für Gymnasien besteht darin, dass beurteilende Statistik auch für die Lehrkräfte neu ist. Sie verfügen über keinerlei Erfahrung auf diesem Gebiet, das sich auch in die ungarische Tradition des gymnasialen Unterrichts schwer einordnen lässt. Deshalb ist es das Ziel des Projekts, eine akzeptable Methode zur Einführung von Inferenzstatistik in die Lehrpläne der Gymnasien zu entwickeln, die auch für die Lehrkräfte tauglich ist.

Die hier vorgestellten Experimente basieren auf allgemeinen Vorschlägen, wie man mit dem Unterrichten von Inferenzstatistik an Gymnasien beginnt, mit Anpassungen und Modifikationen, die unserer Sicht von Statistik, insbesondere der Bedeutung des Bayesianischen Ansatzes und der Wichtigkeit der Einbettung von beurteilender Statistik als Methode in empirischer Forschung entsprechen. Wir halten es für wichtig, Ansatz und Ergebnisse der Pilotstudie zu präsentieren, da sie als Grundlage für eine erweiterte Studie über das neue Curriculum-Design dienen werden; es geht um die Entwicklung und Bewertung von Material für Pilotlehrveranstaltungen oder Workshops an Sekundarschulen in Ungarn. Das Ziel dieser Workshops ist zweifach:

- Alle Personen, die am Workshop teilgenommen haben, werden anschließend hinsichtlich ihres Verständnisses des Materials, ihres erworbenen Wissens und ihres Gesamteindrucks befragt.
- Die Lehrkräfte an den Schulen werden eingeladen, diese Sitzungen zu beobachten und anschließend zu ihrem Verständnis der vorgestellten Experimente und zu ihrer Einstellung zur Einführung dieses Studienbereichs in den Lehrplan befragt, um ihren Bedarf an verschiedenen Unterstützungsmaßnahmen für den Unterricht in beurteilender Statistik zu ermitteln.

Nach vorläufigen, informellen Anfragen ist zu erwarten, dass die Lehrkräfte insgesamt eine sehr ablehnende Haltung gegenüber der Einführung des neuen Themengebiets einnehmen werden, da es sich doch sehr von den anderen Bereichen der Mathematik unterscheidet und sie mit dem Thema und den Begriffen und Ideen nicht vertraut sind. Dies macht auch deutlich, dass eine der Hauptaufgaben im Projekt die Gestaltung der Ausbildung im Lehramt sein wird. In diesem Aufsatz behandeln wir das Design von Experimenten zur Einführung von grundlegenden Ideen und Denkweisen sowie von spezifischen Methoden der beurteilenden Statistik für die Lernenden und nicht für die Lehrkräfte. Die Lehramtsausbildung wird Gegenstand weiterer Aufsätze sein.

Unsere Forschungsfrage lautet: Wie findet man die am besten geeigneten Methoden und Probleme, die – unterstützt durch lebendige und engagierte praktische Unterrichtsaktivitäten – die Konzepte der Inferenzstatistik erstmals Lernenden näherbringen? Ein ähnlicher Ansatz wird in Garfield und Ben-Zvi (2008) und in früheren hier zitierten Studien vorgeschlagen, z. B. in der „Top Ten"-Liste der Empfehlungen von Rossman und Chance (1999).

Wir haben mehrere Experimente entworfen und von einem der Co-Autoren (Jánvári) in einer Klasse des Szerb Antal Gimnázium (Klasse 11; Alter 17 Jahre) testen lassen. Die Experimente werden durch ihre Übereinstimmung mit unserem theoretischen Ansatz zur Modellierungswahrscheinlichkeit und dem parallelen Ansatz zur statistischen Inferenz bestimmt. Unsere Arbeit basiert auf den Ideen zum Unterricht in Wahrscheinlichkeit und Statistik von T. Varga und T. Nemetz (siehe Varga 1972, 1983, Nemetz und Kusolitsch 1999). Diese Arbeit steht im Zusammenhang mit dem Forschungsprojekt „Complex Mathematics Education", das seit 2016 läuft.

Varga sieht die Mathematik als etwas sich ständig Entwickelndes an, wobei für diese Entwicklung Intuitionen und Erfahrungen sowie soziale Aktivitäten entscheidend sind. Deswegen ist sein didaktischer Ansatz durch kreative spielerische Aktivitäten, heuristische Methoden sowie die verwendeten Werkzeuge gekennzeichnet. Reichhaltige Aufgaben und Aufgabensysteme sollen die Begriffe sinnvoll erscheinen lassen. Varga setzt auch auf die Diversität der Begriffe und versucht, verschiedenste Zugänge zu nutzen, um die volle Begriffstiefe zu erschließen, indem in den unterschiedlichen Begriffsaspekten das Verbindende gesucht wird. Glücksspiele und allgemein Spiele nehmen bei Varga eine tragende Rolle ein. Um sie herum werden Aufgabensysteme entwickelt. Nemetz und Kusolitsch transferieren diese Tradition in die Stochastikausbildung der Universität.

Eine weitere Ressource ist die Forschung zu einem parallelen Zugang zur klassischen und Bayesianischen Statistik (Vancsó 2006, 2009, 2018), welche Vargas Idee, aus der Unterschiedlichkeit von Begriffen oder Zugängen zu lernen, exemplifiziert. In dieser Forschung wurden die Einstellungen und Fähigkeiten von Lehrkräften und Lernenden in einem „Parallelkurs" aufgebaut, der Methoden aus beiden Zugängen gegeneinander an denselben Aufgaben kontrastiert. Dieser parallele Ansatz half, die grundlegenden Konzepte und Methoden beider großer Schulen zur statistischen Inferenz – wie sie sich aus fachlicher Sicht ergeben – zu verstehen.

Die bekannten Verständnisprobleme haben die Didaktik der Stochastik immer schon gefordert, seit man statistische Inferenz in den Unterricht der Sekundarstufe aufnehmen wollte. Man hat sich – beginnend mit den 1990er Jahren – damit beschäftigt, wie man die einfacheren Methoden der Re-randomisierung und des Bootstrap auszunutzen kann, um auf die statistische Inferenz vorzubereiten (Borovcnik 2019). In der *Informal Inference* (siehe etwa Garfield und Ben-Zvi 2008) versucht man, Inferenz überhaupt auf diese Methoden zu beschränken. In diesem Ansatz spielen Daten eine Schlüsselrolle. Alles, was zur Inferenz benötigt wird, erhält man durch Simulation aus den Daten; Hypothesen in Form von Wahrscheinlichkeitsverteilungen spielen keine Rolle mehr.

Im Gegensatz dazu versuchen wir, die Inferenz durch Einbeziehen der Bayesianischen Sichtweise verständlicher zu machen. Dies steht im Einklang mit Borovcnik (2019), der Mängel der Informellen Inferenz beschreibt und anregt, den Ansatz insgesamt neu zu überdenken. Unser Hintergrund ist auch durch eine Analyse der statistischen Inferenz aus der Sicht des Statistikers durch Vic Barnett geprägt (Barnett 1982), der versucht, durch den Vergleich der Ansätze diese selbst verständlich zu machen.

Die Bayesianische Statistik im Mathematikunterricht wird seit den 1990er Jahren in Deutschland von Wickmann (1990) untersucht. Wickmann war von der Überlegenheit des Bayesianischen Ansatzes überzeugt und wollte die klassische statistische Inferenz durch die Bayesianische Inferenz ersetzen. Riemer (1991) zeigte die Machbarkeit von Wickmanns Ideen im Umfeld realer Schulklassen auf. Bekannt sind auch die Riemer-Würfel, die rein äußerlich eine vom Würfel abweichende Gestalt haben, sodass von Anfang an eine von der Gleichverteilung unterschiedliche Verteilung ins Blickfeld rückt. Wir haben von Wickmann und Riemer viel über Bayesianisches Denken gelernt.

1997 gab es eine heftige Debatte im Teacher's Corner der Zeitschrift *The American Statistician:* Lindley (1997) und Albert (1997a, b) stellten ihre Ideen über einen Bayes-Zugang vor und verteidigten den Zugang als dem klassischen Testen überlegen (wie Wickmann). Moore (1997a, b) vertrat die Gegenposition. Als es dann dazu kam, die Argumente und die Schlüssigkeit der Zugänge zur statistischen Inferenz auf klassische und auf Bayesianische Art zu bewerten, beendete Moore die Diskussion abrupt und zwar in dem Sinn, dass zwar einige Argumente Bayes favorisieren, aber der Zugang einfach zu schwer sei für die Studierenden.

Der Bayes-Ansatz ist von Borovcnik in den 1980er Jahren als formale Erweiterung des klassischen Ansatzes untersucht worden. Die Leitlinie war, aus der a priori-Verteilung, die nötig wird, um den klassischen Ansatz zu reproduzieren, über die Rationalität des klassischen Verfahrens etwas zu lernen. Es folgte der Versuch, einen gemeinsamen Ansatz von klassisch und Bayesianisch zu entwickeln, was aber wegen der großen Interpretationsunterschiede in der Wahrscheinlichkeitsauffassung von subjektivistisch und frequentistisch zum Scheitern verurteilt war.

In den 1990er Jahren haben mehrere Autoren versucht, die Komplexität der Begriffe beim klassischen Testen durch Anbindung an den Satz von Bayes didaktisch abzufangen. Götz (1997, 2001) analysierte Stochastikaufgaben im Hinblick auf klassische und Bayesianische Behandlung. Wolpers und Götz (2002) listen mehrere Möglichkeiten, Stochastik zu unterrichten. Meyfarth (2009, S. 7) zitiert daraus. Wir beziehen uns nur auf Bayes:

> „Ein an der *Bayes-Statistik* orientiertes Vorgehen, welches dem Satz von Bayes und dem induktiven Schließen beim Hypothesentest einen großen Stellenwert einräumt. Der Ansatz modelliert das „Lernen aus Erfahrung“ und knüpft damit an intuitive Vorstellungen an. Für einen solchen „Bayesianischen Stochastikunterricht“ existiert noch kein geschlossener Lehrgang […].“

Diese Arbeiten zielen darauf ab zu zeigen, dass beim klassischen Zugang zum Testen gewisse Begriffe fehlen, welche die Qualität eines Tests erst erfassen. So kann man beim Testen wohl den α- und β-Fehler angeben (bedingte Wahrscheinlichkeiten, die Hypothese abzulehnen, wenn man unterstellt, dass sie zutrifft, bzw. die Hypothese nicht abzulehnen, wenn man unterstellt, dass eine bestimmte Verteilung aus der Gegenhypothese zutrifft). Allerdings macht die bedingte Wahrscheinlichkeit, dass die Hypothese zutrifft oder eben nicht zutrifft, wenn man Daten erhalten hat, die zur Entscheidung „Ablehnung der Nullhypothese“ führen, im klassischen Ansatz überhaupt keinen Sinn, weil Hypothesen keine Wahrscheinlichkeit zukommen kann. Für die Praxis sind das aber geradezu die entscheidenden Kriterien, ob ein statistischer Test eine gute Qualität hat oder nicht. Es geht hierbei um die korrekte oder falsche Klassifikation. Das kann man etwa im Kontext der Diagnose einer Krankheit mit der Bayes-Formel sehr schön aufzeigen.

Dies war der Kern der Auseinandersetzung im *American Statistician* 1997, dass der klassische Test die wichtigen Fragen der Praxis gar nicht beantworten kann. In diesen didaktischen Arbeiten ging es aber nicht um eine gleichwertige, parallele Entwicklung der klassischen und Bayesianischen Testtheorie, sondern vielmehr darum, aus dem Rahmen der Bayes-Formel aufzuzeigen, wie die Begriffe beim klassischen Testen zu verstehen sind und dass entscheidende Begriffe fehlen und damit wichtige Eigenschaften des Tests nicht bewertet werden können.

In seiner Dissertation schrieb Vancsó (2006) über den Bayesianischen Ansatz in der Schule und entwickelte Ideen, die beiden Ansätze im Unterricht *gleichwertig* zu behandeln, d. h., diese beiden verschiedenen Arten der Inferenzstatistik parallel zu unterrichten. Das schließt auch die Thematisierung des Wahrscheinlichkeitsbegriffs (frequentistisch-objektiv und subjektivistisch-qualitativ) mit ein; beide so unvereinbaren Aspekte von Wahrscheinlichkeit machen erst den vollen Wahrscheinlichkeitsbegriff aus. Seine Experimente mit Lehrkräften wurden in Vancsó (2009) zusammengefasst und in Vancsó (2018) weiterentwickelt.

Die Idee einer parallelen Einführung der beiden divergierenden Wege der Inferenzstatistik in der Schulmathematik entspricht auch Vargas Ideal, die unterschiedlichsten Ansätze zu verbinden. Vancsó greift damit auch Ideen von Mignon und Gamerman (1999) auf, die für die universitäre Ausbildung eine ähnliche Position vertreten haben. Die wesentlichen Bestandteile des Bayesianischen Ansatzes sind die bedingte Wahrscheinlichkeit und der Begriff der subjektiven Wahrscheinlichkeit, der für Mathematiker und Forscher, die stark an die „Objektivität“ der Wissenschaften „glauben“, unangenehm erscheint.

2 Experiment 1: Spiel mit gezinkten Würfeln

Ziel des Spiels ist es, das Grundkonzept, die Begriffe, die Logik sowie die Methode von Hypothesentests durch praktische Erfahrung zu ergründen. Obwohl diese Methode bereits in früheren Arbeiten (etwa Lawton 2009, Dambolena, Eriksen und Kopcso 2009) eingeführt wurde, legt unser Ansatz wesentlich mehr Wert auf das praktische Experiment im Unterricht. Ein spielerischer Zugang zu den Problemen soll die Begriffe aus dem Kontext heraus in natürlicher Weise verständlich machen. Ein weiterer Unterschied besteht darin, dass wir, obwohl das hier nicht näher erläutert wird, die Diskussion auf die alternative Lösung des Problems auf Grundlage Bayesianischer Methoden ausdehnen. Das Würfelexperiment lehnt sich an Riemer (1988) an.

Die einzelnen Schritte im Spiel

Die Lernenden werden in kleine Gruppen eingeteilt; jeder Gruppe wird nach dem Zufallsprinzip ein bestimmter Würfel zugeordnet (siehe Abbildung 1). Es gibt drei Arten von Würfeln, die im Spiel verwendet werden: reguläre, leicht verzerrte und stark verzerrte Würfel. Eine besondere Eigenheit, die durchwegs gut angekommen ist, war, dass wir unsere Würfel im 3D-Drucker selbst hergestellt haben. Jede Gruppe wirft ihren Würfel 50 Mal, zeichnet die Ergebnisse auf und stellt die Daten übersichtlich in einer Häufigkeitstabelle dar.

Abb. 1: Die verwendeten Würfel sind teilweise im 3D-Drucker angefertigt worden

Schritt 0

Die Würfel werden an die Gruppen übergeben mit dem Auftrag, sie in die Hand zu nehmen, zu gewichten und nach Auffälligkeiten zu untersuchen. Die Ansichten über die Würfel sollten untereinander ausgetauscht werden, insbesondere soll diskutiert werden, ob man sie für regelmäßig oder verzerrt hält. Dies kann durch eine subjektive Wahrscheinlichkeit auf den einzelnen Flächen ausgedrückt werden (dies könnte durch eine Wahrscheinlichkeit ergänzt werden, dass diese „Hypothese“ zutrifft, z. B. 0,80). Wenn keine physikalischen Abweichungen gefunden werden, könnten die Lernenden zum „Schluss“ kommen, dass alle Flächen dieselbe Wahrscheinlichkeit haben. Dies könnte auch die Aussage über den Würfel unter Verwendung von „Nullinformation“ sein, die durch die gleichmäßige Verteilung ausgedrückt wird.

Eine solche Zusammenfassung der Information über den Würfel, bevor er geworfen wird, wird als „a priori“-Verteilung auf der Menge der möglichen Ergebnisse des Würfels bezeichnet.

Ziel dieses Experiments ist es auch, die Lernenden darauf hinzuweisen, dass die Eigenschaften des Würfels mit unterschiedlichen Methoden untersucht werden können, nicht allein durch wiederholtes Werfen und Feststellen der Ergebnisse. Zum Beispiel könnten sie den Würfel drehen und fühlen, ob er sich unrund bewegt; sie könnten ihn ins Wasser werfen und überprüfen, ob und wie er schwimmt oder andere physikalische Experimente mit dem Würfel durchführen, die darauf hindeuten könnten, ob die Schwerpunktlinien durch den geometrischen Mittelpunkt verlaufen oder nicht.

Schritt 1

Ziel des Experiments ist es herauszufinden, ob der Würfel regulär (fair) ist oder nicht. Dazu kann der Würfel auch mehrfach geworfen werden. Wie viele Würfe wird man durchführen müssen, um eine ordentliche Bewertung der Ausgangsfrage zu ermöglichen? Wenn man sich auf 50 Würfe einigt, dann soll der Würfel geworfen sowie die Ergebnisse entsprechend aufgezeichnet werden. Die grundlegende Annahme ist, dass der Würfel regulär ist. Die Diskussion sollte sich auch darum drehen, warum man dies voraussetzen muss. Auch ist zu thematisieren, welche Häufigkeiten für die Augenzahlen unter dieser Annahme zu erwarten sind und welche Abweichungen toleriert werden können.

Schritt 2

Der Würfel wird geworfen, die Ergebnisse werden in einer Tabelle zusammengefasst und in einem Diagramm dargestellt (siehe Abbildungen 2 und 3).

Schritt 3

Sobald die Lernenden das Experiment durchgeführt haben, sollten sie ihre Ergebnisse in einem Stabdiagramm darstellen und dann eine erste Annahme treffen, ob sie nun denken, dass ihr Würfel regelmäßig ist oder nicht; diese Annahme soll ausschließlich auf ihren eigenen Daten (und nicht auf den Daten der anderen Gruppen) basieren. Nachdem sie zu einer Entscheidung auf der Grundlage ihrer eigenen Ergebnisse gekommen sind, werden die Ergebnisse mit den anderen Gruppen ausgetauscht und mit ihren eigenen verglichen; daraufhin können die Gruppen ihre Entscheidung über die Eigenschaften ihres eigenen Würfels ändern. In dieser Phase sollte man die Lernenden dazu ermutigen, *ihre* Argumente frei zu äußern, und nur bei Bedarf eingreifen.

Es geht auch darum, sich auf ein geeignetes Maß der Abweichung der Daten (der beobachteten Häufigkeiten für die Augenzahlen) von den erwarteten Häufigkeiten zu einigen. Insbesondere geht es um mathematische Techniken, Unterschiede zwischen zwei „Vektoren“ (die beobachteten und die erwarteten Häufigkeiten der Augenzahlen) zu erfassen: Euklidischer Abstand (Wurzel aus der Summe quadratischer Differenzen der Komponenten), Summe quadratischer Differenzen, Summe absoluter Differenzen oder irgendein anderes Maß?

Schritt 4

Die Gruppen werden gebeten, ihre endgültige Entscheidung über ihren Würfel zu treffen. Es ist eine zentrale Aufgabe schon in dieser Phase, auch die möglichen Konsequenzen der Entscheidung zu diskutieren, was zu einer Diskussion über α- und β-Fehler im Kontext führt. Als Abschluss des Spiels könnten wir auch diejenigen „belohnen“, welche die richtige Entscheidung getroffen haben, und diejenigen, die es nicht getan haben, „bestrafen“.

Außerdem können wir jene, die einen α-Fehler in ihrer Entscheidung gemacht haben (die Nullhypothese H_0 – die Würfel sind regulär – ablehnen, obwohl sie zutrifft) und diejenigen, die einen β-Fehler gemacht haben (die Nullhypothese H_0 akzeptieren, obwohl der Würfel verzerrt ist und sie daher nicht zutrifft), anders „bestrafen“. In der Denkweise, wie Tests in der Praxis verwendet werden, führen α-Fehler häufiger zu größeren Problemen, weshalb diese Art von Fehlern stärker „bestraft“ werden sollte. Dies etwa deshalb, weil die Nullhypothese sozusagen als üblicher Zustand, als Normalzustand oder als „Business as usual“ angedacht ist und man sich nur dann dagegen entscheiden sollte, wenn man durch die Daten ziemlich davon überzeugt wird, dass es dieses Mal anders sein könnte.

Die gleichen Aufgaben können auch mit einem Zufallszahlengenerator mit der entsprechenden Spezifikation durchgeführt werden. Wenn wir die Möglichkeit haben, Computer mit der gleichen Gruppe von Lernenden zu benutzen, können wir das gleiche Problem mit viel größeren Datensätzen untersuchen.

Die Software ermöglicht es auch zu beobachten, wie sich die Verteilung mit bis zu 500 oder sogar 5000 Würfen ändert. Die Simulation selbst stellt natürlich einen Bruch des direkt experimentellen Zugangs dar. Sie wird jedoch in ihrer Wirkweise verstanden und akzeptiert, wenn sie das tatsächliche Spiel mit dem didaktischen Gerät übernimmt. Auf diese Weise können die Lernenden erkennen, dass die Erhöhung des Stichprobenumfangs die Entscheidungsfindung einfacher und zuverlässiger macht, d. h. wie sowohl α- als auch β-Fehler reduziert werden können.

Die Ergebnisse aus dem Unterricht, etwas besser dargestellt, sind in den Abbildungen 2 und 3 enthalten. Wie Abbildung 2 zeigt, ähnelt das Stabdiagramm bei 50 (Excel-simulierten) Würfen eines regulären Würfels nicht unbedingt einer gleichmäßigen Verteilung – selbst bei 500 Würfen könnte es ungleich aussehen. Das 5000fache Werfen scheint jedoch die erwarteten Ergebnisse zu liefern.

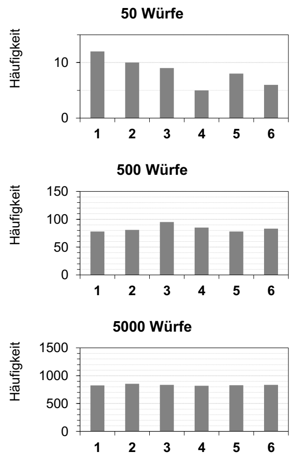

Abb. 2: Ergebnisse eines fairen Würfels, der 50, 500 bzw. 5000 Mal geworfen wurde

Man beachte, dass die *verschiedenen Skalen in Abbildung 2 das Auftreten von relativen statt von absoluten Häufigkeiten nachahmen*, weil jeweils derselbe Anteil an der vollen Spannweite in der Zeichnung gleich groß dargestellt wird. Es wird ja in einem ersten Ansatz gern die Verwendung von relativen Häufigkeiten vermieden, was zu einiger Verwirrung führen kann, weil natürlich die absoluten Häufigkeiten mit Zunahme der Länge des Experimentes divergieren, während man doch die Konvergenz der relativen Häufigkeiten ausnutzen möchte, damit man belegt, dass längere Serien die unbekannten Wahrscheinlichkeiten zuverlässiger einschätzen lassen (siehe auch Borovcnik 2019).

Eine weitere Möglichkeit besteht darin, Simulationen für unterschiedlich verzerrte Würfel durchzuführen. Unterstützt mit Stabdiagrammen kann man untersuchen, wie viele Würfe man benötigt, um gute Entscheidungen zu treffen. Wie Abbildung 3 zeigt, ist ein stark verzerrter Würfel schon nach 50 Würfen zu erkennen. Andererseits kann *weder* ein regulärer *noch* ein leicht verzerrter Würfel „bestätigt“ werden

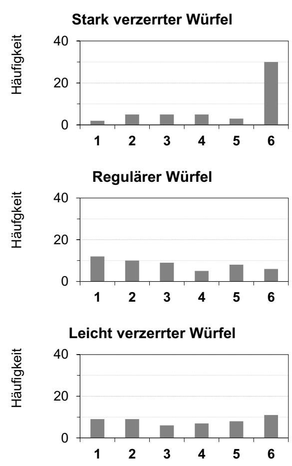

Abb. 3: Ergebnisse von 50 Würfen mit verschiedenen, auch gezinkten Würfeln

(das hat mit Testlogik und β-Fehlern zu tun). Basierend auf Daten könnten wir falsche Entscheidungen über reguläre oder mäßig verzerrte Würfel treffen:

Für den regulären Würfel könnten wir fälschlich feststellen, dass es sich um einen verzerrten Würfel handelt und damit einen α-Fehler begehen, während wir für einen mäßig verzerrten Würfel fälschlicherweise feststellen könnten, dass der Würfel regulär ist, und damit einen β-Fehler begehen.

Didaktische Ziele des Experiments

Das Experiment unterstellt eine Wahrscheinlichkeit für jede Augenzahl. Das ist die Grundannahme eines jeden stochastischen Experiments.

Die Würfel wurden erst nur 50 Mal geworfen, dann insgesamt 200 Mal. Das mag als viel zu gering erscheinen. Wir haben mit Absicht diese kleinen Serien gewählt. Die Variabilität der Ergebnisse sollte noch so groß sein, dass eine Entscheidung über den Würfel schwierig wird. Wenn man sehr oft wirft (mindestens 1000 Mal), so werden die relativen Häufigkeiten sehr nahe bei den Wahrscheinlichkeiten liegen und es wird leicht, Abweichungen von einem idealen Würfel zu erkennen. Außerdem geht es ja auch um das *Gefühl* für die Grenze, die tolerabel ist. Wir haben ja keinerlei Erfahrung mit langen Wurfserien. Für kürzere Serien sollten wir aber „spüren“, welche Abweichungen mit einem idealen Würfel vereinbar sind und welche nicht.

Die Testlogik lässt nicht zu, dass eine Hypothese bestätigt wird. Wenn die Nullhypothese nicht abgelehnt werden konnte, haben wir auch andere Hypothesen untersucht. Diese konnten dann auch nicht abgelehnt werden. Das war ein deutliches Zeichen, dass die Methode des statistischen Tests einer Verwerfungslogik unterliegt: Bleib so lange bei der Nullhypothese, wie die Daten damit kompatibel sind. Verwirf die Nullhypothese, wenn die Daten deutlich dagegen sprechen.

Schritt 3 ist unerlässlich, aber er stellt die Lernenden vor die Herausforderung, geeignete Kriterien für ihre Entscheidung zu finden. Die geführte Diskussion über Möglichkeiten der Differenzmessung bereitet die Einführung des Chi-Quadrat-Tests vor.

In Schritt 4, bei der Entscheidungsfindung und der Analyse möglicher Fehler, die begangen werden können, machen sich die Lernenden mit der Bedeutung und Relevanz von α- und β-Fehlern vertraut. Wir müssen nicht alle diese Begriffe im Unterricht benennen und können sie durch Analogie mit einem (Bayesianischen) Entscheidungsproblem besser einschätzen (siehe z. B. Wolpers und Götz 2002). Es muss betont werden, dass alles, was wir hier sagen können, Wahrscheinlichkeiten betrifft: Nichts kann mit Sicherheit ausgedrückt werden; es gibt immer eine Wahrscheinlichkeit ungleich Null für Fehler. Am Ende können wir einen (χ^2-)Anpassungstest durchführen. Durch dieses Experiment werden neue Methoden zur Prüfung der *unterstellten* Verteilung des Würfels im Hinblick auf die Datenlage eingeführt. Diskussion und Gedankenaustausch bieten die Möglichkeit, *gemeinsam* zu denken. Die Aktivität bereichert die Lektion aus methodischer Sicht.

Es zeigt sich jedoch, dass man die Problematik des Signifikanztests besser verstehen kann, wenn man die Bayes-Analyse einbezieht. Das kam aus dem Klassengespräch heraus, welches illustriert, wie die Lernenden mit dem Verständnis der Vorgangsweise ringen. Siehe dazu Experiment 2.

3 Experiment 2: Würfeln mit Bayes

Dieses Experiment wurde noch nicht explizit durchgeführt, es hat sich jedoch im Rahmen von Experiment 1 erwiesen, dass man auf Bayesianische Aspekte eingehen soll, wenn man den Signifikanztest verständlich machen möchte. Wir hatten den Bayes-Ansatz im Design von Experiment 1 vorerst zurückgestellt aus der Überlegung heraus, im ersten Ansatz technische Details aus dem Spiel zu lassen. Unser Plan war es, direkt in die Problematik des Signifikanztests am Beispiel des χ^2-Tests einzuführen.

Erkenntnisse aus Experiment 1

Im ersten Experiment wurden die Würfel auch physisch (visuell und haptisch, am Ende auch durch ein physikalisches Experiment) auf etwaige Abweichungen von einer Gleichverteilung hin untersucht. Es wurden Vermutungen über das Auftreten der einzelnen Augenzahlen (deren Wahrscheinlichkeiten) bei den verschiedenen Würfeln geäußert.

In der Diskussion um die Relevanz und die Interpretation der χ^2-Ergebnisse kam die Frage wieder hoch: Die Lernenden wollten wissen, was der χ^2-Test nun für die Wahrscheinlichkeiten des Würfels bedeutet.

> „Warum haben wir den Würfel untersucht und zuvor eine Stunde Physik gemacht, wenn wir unsere Einschätzung für die einzelnen Augenzahlen gar nicht verwenden?“

Diese Frage stand im Raum. Es gab speziell zwei Schüler, die so argumentierten:

> „Wenn wir Abweichungen von der Gleichverteilung erkennen, warum benutzen wir dann einen Test, anstatt dass wir diese Abweichungen zum Weiterrechnen verwenden?“

Die Frage war auch: Und wenn wir bei der physikalischen Untersuchung keine Abweichungen finden, warum verwenden wir die Annahme, dass keine Abweichung besteht und testen noch einmal, dass keine Abweichungen bestehen? Die Lernenden hatten sich damit in der Logik des Testens „verbissen".

Wir sind im Klassengespräch diesen Fragen so begegnet: Man muss die physikalische und die statistische Prüfung des Würfels voneinander unterscheiden. Statistisch wird der Würfel lediglich über die Wurfergebnisse geprüft. Die statistische Prüfung wird im Gegensatz zur Prüfung der physikalischen Eigenschaften auch empirische Evidenz genannt.

Dazu führen wir den χ^2-Test ein und zeigen, wie er funktioniert. Weil er nur auf Daten basiert, werden die physikalischen Eigenschaften des Würfels ignoriert. Die zweite Frage, wenn wir keine Abweichungen feststellen, ist leichter zu beantworten: wir könnten etwas übersehen haben, denn die Ursachen für Abweichungen können sehr verborgen sein.

Die Vorgangsweise beim Bayes-Ansatz

An dieser Stelle begnügten wir uns damit, die Lernenden qualitativ darauf hinzuweisen, wie man diese Wahrscheinlichkeiten (die ja subjektiv geprägt sind) verwenden kann. Wir illustrierten anhand eines Kontexts (Lotto), was die Problematik ist und wie man vorgehen könnte.

Wir erklärten die Vorgangsweise mit Skizzen der durch die Daten revidierten Verteilungen und verwiesen darauf, dass es entsprechende (nur beim ersten Hinsehen komplizierte) Algorithmen gibt, welche diese Berechnungen übernehmen. Wir würden dann sozusagen nur die Interpretation der jeweiligen Verteilungen vornehmen, welche unsere a priori-Einschätzung der Wahrscheinlichkeiten des Würfels und die Daten berücksichtigen (siehe Experiment 3). Es wurde auch angesprochen, dass die entstehenden Verteilungen unsere aktuelle Einschätzung für die Augenzahlen wiedergeben und dass diese Wahrscheinlichkeiten durch die zunehmende Zahl der Daten objektiver werden.

Basierend auf den *a priori*-Wahrscheinlichkeiten, die in Schritt 0 des Spiels in Experiment 1 geschätzt wurden (*bevor* man Daten hat), erhalten wir mit dem Bayes-Satz eine neue Verteilung bezüglich der sechs Ergebnisse des Würfels, welche auch die Daten berücksichtigt. Dieser neue Kenntnisstand (*nachdem* man Daten hat) wird *a posteriori*-Verteilung genannt. Sie wird sich bei den Lernenden aufgrund ihrer subjektiven Vorverteilung unterscheiden. Durch die Erhöhung der Anzahl der Experimente jedoch konvergiert die Folge der a posteriori-Verteilungen, die entsteht, wenn man sukzessive Daten bekommt und diese zum Revidieren der Verteilung verwendet, zur „wahren", zur tatsächlichen Verteilung des regulären oder verzerrten Würfels; das passiert zufolge des zentralen Satzes im Bayesianischen Ansatz (siehe etwa Wickmann 1990). Die Grenzverteilung ist unabhängig von der gewählten (und damit subjektiven) a priori-Verteilung.

Der Bayesianische Ansatz arbeitet mit einer a priori-Verteilung der Wahrscheinlichkeiten der sechs verschiedenen Ergebnisse, die durch die Wahrscheinlichkeitsfunktion $f_{\text{priori}}\ (x_1, x_2, x_3, x_4, x_5, x_6)$ ausgedrückt wird. Dann erhalten wir ein Ergebnis der Experimente (durch Werfen des Würfels). Dadurch entsteht ein neue Verteilung $f_{\text{posteriori}}\ (x_1, x_2, x_3, x_4, x_5, x_6)$. Die Technik erscheint kompliziert, aber wir können die a posteriori-Verteilung per Software bestimmen. Es bleibt, den Mechanismus der Berechnung zu erklären. Die a priori-Verteilung fasst unser Wissen über den Würfel zusammen, bevor wir ihn werfen. Die a posteriori-Verteilung kombiniert dieses Wissen mit dem Wissen aus den Daten (was der Wurf ergeben hat).

Dieser Ansatz unterscheidet sich vom klassischen Modell. Insbesondere bekommen Hypothesen auch eine Wahrscheinlichkeit zugeordnet; folglich kann man auch von bedingten Wahrscheinlichkeiten von Hypothesen sprechen, wobei als Bedingung die vorhandenen Daten eingesetzt werden. Über die Philosophie dahinter sei etwa auf Gigerenzer (1993) verwiesen. Auf dem klassischen Weg in Experiment 1 nähern wir uns den Begriffen Null- und Alternativhypothese aus der von Neyman und Pearson entwickelten klassischen Testtheorie (Neyman 1952). Historisch gesehen hat R. A. Fisher dieses Problem zuerst angegangen (Fisher 1935), aber er hat nur das Konzept des α-Fehlers eingeführt (als Diskrepanzmaß) und hat sich dagegen verwehrt, dass man zur statistischen Beurteilung auch eine Alternativhypothese benötigt.

Es ist wichtig zu betonen – um den Ansatz richtig zu begründen –, dass die Nullhypothese eine spezifische Aussage (eine bestimmte Verteilung) ist, während die alternative Hypothese die Nichterfüllung dieser Aussage ist und daher *alle anderen (im Sachverhalt plausiblen) Verteilungen* der Modellfamilie umfasst. Das macht Wahrscheinlichkeitsberechnungen unter der Annahme der Alternativhypothese schwieriger, weil der β-Fehler zu einer Funktion wird. Man muss hier so etwas wie einen *Worst Case* für einen Parameter aus der Alternative finden und Argumente vorbringen, warum es reicht, die Analysen nur auf diesen

Fall zu beschränken, damit man die Komplexität von β wieder reduziert.

4 Experiment 3: Anzahl der Kugeln im Lotto

Das folgende Experiment (Vancsó 2009) diente bei der Besprechung von Experiment 1 und den Schwierigkeiten mit dem Verständnis des χ^2-Tests eine illustrierende Rolle. Das Experiment wurde mehrfach im Unterricht erprobt; im gegenwärtigen Unterricht spielte es nur eine illustrative Rolle.

Es geht um Lotto. In Deutschland werden 6 aus 49 Kugeln ohne Zurücklegen gezogen. Der Kontext unterstellt, dass man in ein fernes Land kommt und die Ergebnisse einer Ziehung sieht: 4, 6, 15, 19, 40, 37. „Man weiß man aber noch nichts über die Zahl *N* der Kugeln in der Lottotrommel.“ Die anfängliche Unkenntnis – modelliert mit einer Gleichverteilung auf dem Intervall [30, 80] für die Anzahl der Kugeln *N* oder subjektiv als Gleichverteilung empfunden – wird dann Woche für Woche durch die gezogenen Zahlen verändert und man erhält rasch eine sehr scharfe Information, weil sich die revidierte Verteilung für *N* schon nach wenigen Wochen auf ein bis zwei Werte zusammenzieht. Damit führt die anfänglich subjektive Gleichverteilung für *N* sukzessive zu einer objektiveren Einschätzung. Wir zeigen nur die Ausgangsverteilung (oben in Abbildung 4) und drei Zwischenstände, die dem Informationsstand entsprechen, wenn man die Lotto-Zahlen aus 1, 5 bzw. 10 Wochen kennt (unten in Abbildung 4).

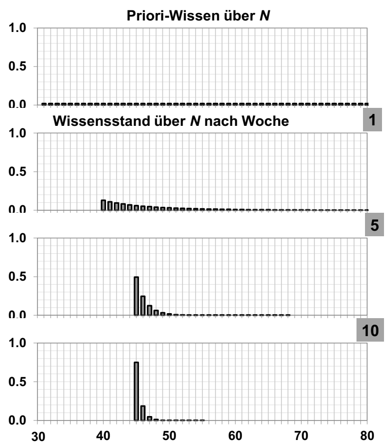

Abb. 4: Der ursprüngliche Wissensstand über die Zahl *N* der Kugeln in der Urne wurde durch eine Gleichverteilung modelliert (a priori-Verteilung, oben). Nach jeder Woche mit den jeweils gezogenen Zahlen wird der Kenntnisstand angepasst (und durch die a posteriori-Verteilung ausgedrückt). Die Diagramme darunter spiegeln den aktuellen Kenntnisstand nach den Ergebnissen von 1, 5 bzw. 10 Wochen.

5 Experiment 4: Stopp die Zeit

Alle Lernenden erhalten eine Stoppuhr und müssen diese nach einer bestimmten Zeit anhalten. Wir können eine manuelle Stoppuhr oder einen digitalen Stopper auf Smartphones verwenden. Das Experiment sollte gleichzeitig durchgeführt werden; zunächst wird nur das Ergebnis des *ersten* Versuchs aufgezeichnet. Dann sollen sie den Versuch wiederholen (fünf Mal etwa) und den letzten Versuch aufzeichnen. Während der erste Versuch das vorhandene Zeitgefühl repräsentiert, wird im letzten Versuch ein etwaiger Lernfortschritt eine Rolle spielen.

Ein weiteres Klassenexperiment beschäftigt sich mit dem subjektiven Zeitgefühl aus einer anderen Sicht. Die Lernenden zählen 60 Sekunden und vergleichen das Ergebnis mit der tatsächlichen Zeit, die von einer Stoppuhr angezeigt wird. In einem ersten Versuch können sie auf die Stoppuhr schauen; in den Wiederholungen des Experiments ist es ihnen nicht erlaubt, die Uhr zu betrachten. Im Unterricht wird man eine kürzere Zeitspanne wählen (etwa 5 Sekunden; siehe Abbildung 5).

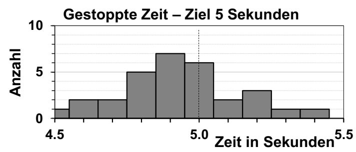

Abb. 5: Daten eines Experiments mit der Stoppuhr

Experimente dieser Art eignen sich gut zur Hinführung zu Hypothesentests: Man kann die Ausgangsdaten zur biologischen Uhr mit den Daten zum Lernerfolg vergleichen und prüfen, ob ein Lernerfolg vorhanden ist oder nicht. Desgleichen kann man prüfen, ob die Ergebnisse von Schülern mit denen von Schülerinnen vergleichbar sind. Wer hat also das genauere Zeitgefühl?

Das führt zu den üblichen Fragestellungen der empirischen Forschung und methodisch zum Zwei-Stichproben-t-Test. Während die Verteilungen für die Sekundarstufe außer Reichweite stehen, kann man die Methode auf informeller Basis durch geeignete Simulationsszenarios gut motivieren (siehe auch Borovcnik 2019).

6 Schlussfolgerungen

Unser erstes Experiment, das Würfeln, wird verwendet, um Hypothesentests in der Mathematik der Sekundarschule einzuführen. Der Bayesianische Weg (Experiment 2) kann – obwohl zunächst kompliziert – mit geeigneter Software durchaus überschaubar gehalten werden. Das hat auch schon Wickmann (1990) erkannt und eine eigene Software geschrieben. Es geht nicht um technische Details oder darum, diese auch durchzurechnen, sondern es geht vielmehr um ein qualitatives Verständnis von Wahrscheinlichkeitsverteilung als Ausdruck eines Urteils oder eines Informationsstandes über gewisse Möglichkeiten und wie diese Verteilung durch weitere Daten, die man erhält, verändert wird (siehe auch Vancsó 2018). Experiment 3 mit der unbekannten Zahl der Lotto-Kugeln diente zur Illustration der Vorgangweise beim Bayesianischen Zugang zum Testen; insbesondere wird thematisiert, wie sich ein subjektiver Wissensstand durch zunehmende Daten objektiviert und wie man das Wissen durch eine Verteilung beschreibt.

Für die Hauptphase der Unterrichtsexperimente werden wir Aktivitäten vorbereiten, welche den Bayes-Ansatz beim Würfeln verfolgen. Die Herausforderung wird sein, den Lernenden mit Analogien und Meta-Wissen beizustehen, um die technischen Anforderungen in der Qualität zu verstehen, die Berechnungen aber der Software zu überlassen.

Experiment 4 verbindet Messtechnik und physikalische Fragen der Messgenauigkeit mit statistischen Fragen der Verteilung und der Beschreibung dieser Verteilung. Der Kontext der biologischen Uhr leitet in zwangloser Weise auf Fragen über, die man mit statistischen Tests beantworten kann. Das Experiment eignet sich auch dazu, den Begriff der Kontinuität und von stetigen Verteilungen zu thematisieren. Diese Verwendung im Unterricht wird in einem weiteren Aufsatz beschrieben werden.

Die Pilotstudie im Unterricht hat die Durchführbarkeit der Experimente gezeigt. Das Feedback der Lernenden (drei Sätze darüber, was dir gefallen hat und was dir an dem Experiment am Ende der Experimentalklasse nicht gefallen hat) ist ermutigend, unseren Ansatz fortzusetzen:

> „Wir wollen nicht aufhören“, „[…] ist das das Diagramm, das wir bekommen sollen“, „Ich mag es, dass wir im Mathematikunterricht experimentieren; das machen wir normalerweise nicht.“

Das Hauptexperiment wird von einer Dokumentation sowie einem Test am Ende des Unterrichts mit geeigneten Beispielen begleitet werden, damit wir die Effizienz des Ansatzes belegen können.

Es ist unser Ziel, mehr Experimente und Spiele im Unterricht einzusetzen, was für die Lernenden nicht nur motivierend ist, sondern auch eine ausgezeichnete Gelegenheit bietet, grundlegende Begriffe und Missverständnisse in Wahrscheinlichkeit und Statistik zu diskutieren. Dieses Ziel manifestiert auch die von Székely (1986) in seinem berühmten Buch über Paradoxien entwickelte Perspektive.

Auch der Einsatz moderner digitaler Werkzeuge kann im Unterricht am Gymnasium motivierend sein; dazu gehören selbstverständlich auch Mobiltelefone. Dies leitet auf den Einsatz von Algorithmen über und spricht eigentlich noch mehr für die Einbindung des Bayes-Ansatzes und eine ingenieurmäßige Herangehensweise, welche üblicherweise die mathematischen Lücken mit Analogien und Meta-Wissen füllt und die Werkzeuge durch wiederholten Gebrauch als nützlich erscheinen lässt.

Die Visualisierung und ein allgemeineres Konzept von Wahrscheinlichkeitsverteilung spielen hierbei eine tragende Rolle. Visualisierung trägt zu einem breiteren Verständnis von Wahrscheinlichkeitsverteilungen bei, die in diesem Ansatz oft sehr stark von einer Verteilung um ein Zentrum abweichen und daher nicht mehr so einfach durch einen Lage- und einen Streuungsparameter beschrieben werden können. Subjektive und objektive Aspekte in einer Gesamtschau ergeben erst ein volles Bild des Begriffs Wahrscheinlichkeit. Hier sei auf Steinbring (1991) verwiesen, der von komplementären Aspekten und vom theoretischen Charakter von Wahrscheinlichkeit spricht, wobei der Begriff nachteilig verändert wird, wenn man sich in unterrichtlichen Aktivitäten auf einen Teil beschränkt (er bezieht sich hierbei auf relative Häufigkeit und Anteil).

Varga (1983) spricht die Bedeutung von Intuitionen und Erfahrungen für die Mathematik und fürs Lernen von Mathematik an. Heuristische Methoden sollen die mathematischen Begriffe vorbereiten. Werkzeuge und Spielzeuge – geeignet genutzt – sollen durch kreative spielerische Aktivitäten nicht nur die Freude an der Mathematik wecken sondern in angeleiteter und dennoch aktiver Entdeckung zu tieferem Verständnis der Begriffe führen.

Experimente als Aufgabensysteme in einen inneren Zusammenhang zu bringen, war Varga ein wesentliches didaktisches Anliegen. Dabei spielt auch die Vielfalt an Zugängen eine Rolle: wir haben den Bogen von Würfelexperimenten bis hin zu Experimenten zum subjektiven Zeitempfinden gespannt und zunächst einfach anmutende Aufgaben in Aufgabensysteme weiterentwickelt, welche die vielfältigen Fragen aus Messtechnik, Biologie sowie statistischer Inferenz in natürlicher Weise entwickeln lassen und zum Verständnis der Antworten der Statistik beitragen. Das Gamifikationselement (Papp 2017) kann Engagement wesentlich erhöhen und damit Verständnis für das wissenschaftliche Material ermöglichen. Damit schließt sich der Kreis zur Aussage eines Schülers im Pilotprojekt:

„Es bringt mich dazu, mehr verstehen zu wollen.“

Danksagung

Die Arbeit wurde von der Ungarischen Akademie der Wissenschaften gefördert.

Literatur

Albert, J. (1997a): Teaching Bayes’ rule: A data-oriented approach. In: *American Statistician* 51(3), S. 247–253.

Albert, J. (1997b): Reply. In: *American Statistician* 51(3), S. 271.

Barnett, V. (1982): Comparative Statistical Inference. New York: Wiley.

Batanero, C.; Borovcnik, M. (2016): Statistics and probability in high school. Rotterdam: Sense Publishers

Borovcnik, M. (2019): Informal and ‚informal‘ inference. In: Contreras, J. M.; Gea, M. M.; López-Martín, M. M.; Molina-Portillo, E. (Hrsg.): Actas del Tercer Congreso International Virtual de Educación Estadística (S. 1–17). Granada: Universidad Granada. www.ugr.es/~fqm126/civeest.html (Zugriff: 26.7.2019)

Dambolena, I. G.; Eriksen, S. E.; Kopcso, D. P. (2009): An intuitive introduction to hypothesis testing. In: *INFORMS Transactions on Education* 9(2), S. 53–62.

Fisher, R. A. (1935): The design of experiments (2. Aufl., 1937). Edinburgh: Oliver and Boyd.

Garfield, J. B.; Ben-Zvi, D. (2008): Learning to reason about statistical inference. In: Garfield, J. B.; Ben-Zvi, D. (Hrsg.): Developing students’ statistical reasoning (S. 261–288). New York: Springer.

Gigerenzer, G. (1993): The Superego, the Ego, and the Id in statistical reasoning. In: Keren, G.; Lewis, C. (Hrsg.): A handboook for data analysis in the behavioral sciences: Methodological issues (S. 311–339). Hillsdale, NJ: Lawrence Erlbaum.

Götz, S. (1997): Bayes-Statistik – ein alternativer Zugang zur beurteilenden Statistik in der siebenten und achten Klasse AHS. Dissertation, Univ. Wien.

Götz, S. (2001): Klassische und Bayesianische Behandlung von Stochastikaufgaben aus österreichischen Schulbüchern. In: Borovcnik, M.; Engel; J; Wickmann, D. (Hrsg.): Anregungen zum Stochastikunterricht (S. 147–162). Hildesheim: Franzbecker.

Lawton, L. (2009): An exercise for illustrating the logic of hypothesis testing. In: *Journal of Statistics Education* 17(2).

Lindley, D. V. (1997): Discussion. In: *American Statistician* 51(3), S. 265–266.

Meyfarth, T. (2009): Die Konzeption, Durchführung und Analyse eines simulationsintensiven Einstiegs in das Kurshalbjahr Stochastik der gymnasialen Oberstufe. Kasseler Online-Schriften zur Didaktik der Stochastik. Kassel: Universität Kassel. urn:nbn:de:hebis:34-2006100414792 (Zugriff 2.1.2020)

Mignon, H. S.; Gamerman, D. (1999): Statistical inference: An integrated approach. London: Arnold.

Moore, D. S. (1997a): Bayes for beginners? Some reasons to hesitate. In: *American Statistician* 51(3), S. 254–261.

Moore, D. S. (1997b): Reply. In: *American Statistician* 51(3), S. 272–274.

Nemetz, T.; Kusolitsch, N. (1999): Guide to the Empire of Random. Budapest: TypoTex.

Neyman, J. (1952): Lectures and conferences on mathematical statistics and probability. Washington, DC: Graduate School, U.S. Dept. of Agriculture.

Papp, T. A. (2017): Gamification effects on motivation and learning: Application to primary and college students. In: *International Journal for Cross-Disciplinary Subjects in Education* 8(3), S. 3193–3201.

Riemer, W. (1988): Riemer-Würfel. Stuttgart: Klett.

Riemer, W. (1991): Stochastische Probleme aus elementarer Sicht. Mannheim: Spektrum Verlag.

Rossman, A. J.; Chance, B. L. (1999): Teaching the reasoning of statistical inference: A „top ten“ list. In: *The College Mathematics Journal* 30(4), S. 297–305.

Steinbring, H. (1991): The theoretical nature of probability in the classroom. In: Kapadia, R.; Borovcnik, M. (Hrsg.). Chance encounters (S. 135–167). Dordrecht: Kluwer.

Székely, G. (1986): Paradoxes in probability theory and mathematical statistics. Dordrecht: D. Reidel.

Vancsó, Ö. (2006): Classical and Bayesian statistics from didactical point of view. Dissertation: Univ. Debrecen.

Vancsó, Ö. (2009): Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. In: *International Electronic Journal of Mathematics Education* 4(3), S. 291–322. www.iejme.com/volume-4/issue-3. (Zugriff: 26.7.2019)

Vancsó, Ö. (2018, Juli): How visualisation using software helps understanding classical and Bayesian statistics. *Invited paper „Teaching Probability in School – Understanding and Linking it to Statistics.“* ICOTS 10, Kyoto. www.researchgate.net/profile/Oedoen_Vancso. (Zugriff: 26.7.2019)

Varga, T. (1972): Logic and probability in the lower grades. In: *Educational Studies in Mathematics* 4(3), S. 346–357.

Varga, T. (1983): Statistics in the curriculum for everybody – How young children and how their teachers react. In: Grey, D. R et al. (Hrsg.): Proceedings of the First International Conference on Teaching Statistics Bd. 1 (S. 71–80). Sheffield: Teaching Statistics Trust.

Wickmann, D. (1990): Bayes-Statistik: Einsicht gewinnen und entscheiden bei Unsicherheit. Mannheim: BI.

Wolpers, H.; Götz, S. (2002): Mathematikunterricht in der Sekundarstufe II. Band 3. Didaktik der Stochastik Braunschweig: Vieweg.

Anschrift der Verfasser

Manfred Borovcnik
Institut für Statistik, Universität Klagenfurt
Universitätsstraße 65, 9020 Klagenfurt
manfred.borovcnik@aau.at

Peter Fejes Tóth
Biometrie & Agrarinformatik, Szent-István Egyetem
Villány út 29-43, 1118 Budapest
fejestothpeter@yahoo.com

Zsuzsanna Jánvári
Mathematikdidaktisches Zentrum
Eötvös Loránd Tudományegyetem
Pázmány Péter setány 1/A, 1117 Budapest
zsjanvari@gmail.com

Ödön Vancsó
Eötvös Loránd Tudományegyetem
Pázmány Péter setány 1/A, 1117 Budapest
vancso@caesar.elte.hu