Stochastische Konzepte und Modelle mit dynamischen Applets interaktiv erforschen

MANFRED BOROVCNIK, KLAGENFURT

Wir gehen auf einige Besonderheiten der Stochastik ein, die sie für innovative Lernformen immer schon empfänglich gemacht haben. Neue Medien wurden daher frühzeitig im Stochastik-Unterricht eingesetzt, nicht nur für komplexe Berechnungen. Wir stellen einige didaktisch inspirierte dynamische Visualisierungen vor, welche wir auch in unserer Einführung in die Wahrscheinlichkeitsrechnung und Statistik einsetzen, um das Verständnis für komplexere Konzepte zu unterstützen. Bei der Erstellung der Applets lassen wir uns von folgenden Zielvorstellungen leiten: i) Der abstrakte Begriff Wahrscheinlichkeit wird durch seine Spuren in relativen Häufigkeiten in wiederholten Versuchen illustriert. ii) Die Auswirkung von Parametern in Modellen wird durch systematisches Variieren gezeigt. iii) Zentrale Sätze (Grenzwertsätze) werden oft durch Simulation in Aktion demonstriert, aber anstelle eines – materiell nicht darstellbaren – Grenzwerts wird nur die Monotonie des Verhaltens gezeigt. Durch Erkennen eines stabilen Musters wird ein Gedankenexperiment angeregt, das den Sachverhalt des Satzes verständlich macht. Allen Applets gemeinsam ist ein dynamisches Verändern der Situation. Wie in einem Film sieht man mathematische Sätze oder Sachverhalte im Entstehen.

1. Einleitung

Wir gehen auf Schwierigkeiten mit der Begriffsbildung und dem Begriffserwerb in Stochastik kurz ein. Das hat schon immer die Suche nach neuen Lernformen befeuert; für die Elementarisierung und Visualisierung sind schon früh Neue Technologien und das Internet als Ergänzung für den Unterricht einbezogen worden. Die computerintensiven Methoden der Fachstatistik waren dabei durchaus auch Vorbild für fachdidaktische Ansätze. Blended Learning wird als gute Gelegenheit angesehen, den klassischen Präsenzunterricht zu ergänzen. Wir werden dynamische Applets vorstellen, die sowohl die Berechnungen, Visualisierungen als auch das Begriffsverständnis erleichtern sollen.

Dabei tangieren wir Schlüsselstellen für den Stochastik-Unterricht, und das unabhängig davon, ob wir die Überlegungen auf eine einführende Lehrveranstaltung Statistik für Wirtschaftswissenschaften oder die Stochastik auf Schulniveau beziehen. Diese Studierenden sind mit Schülern vergleichbar; im Hinblick auf ihre Mathematikkenntnisse liegen sie etwas darunter, weil sie in der Regel nach der Schule Mathematik vermeiden wollten. Das Potential der Applets geht aus der exemplarischen Vorstellung hervor; es lohnt sich aber, es eigens zusammenzufassen. Die fachdidaktischen Überlegungen stehen unter dem Motto „Mathematische Einsichten durch Explorieren nachvollziehen". Wir versuchen, die Mathematik auf eine Meta-Ebene zu verlagern und durch Probieren, Simulieren, Visualisieren und Nachverfolgen der Wirkung von Änderungen von Parametern der untersuchten Modelle Muster und mathematische Gesetze aufzuspüren. Excel unterstützt diese Prozesse und ermöglicht interaktives Lernen. Studierende erhalten damit stabile Vorstellungen von Art und Relevanz der Begriffe. Der Aufsatz dient dazu, das Verhältnis von Intuitionen und Mathematik zu stärken, ganz im Sinne von *Stochastik im Wechselspiel von Intuitionen und Mathematik* (Borovcnik 1992).

2. Innovative Lernformen

Wahrscheinlichkeitsrechnung und Statistik gelten seit je her als schwierig. Wir gehen auf ihre Besonderheiten ein und auch darauf, welche didaktischen Antworten versucht wurden. Daran schließt sich eine Betrachtung von Technologie-Einsatz und E-Learning an, welche gerade in der Stochastik verbreitet sind. Blended Learning meint, Technologie im Rahmen traditioneller Vermittlung der Inhalte zu verwenden, so quasi als Ergänzung des Lernens in einer klassenähnlichen Situation. Dazu wird eine Reihe von für den Begriffserwerb kritischen Kategorien diskutiert.

Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), Heft 51, 2018, S. 1-16

2.1 Der Bedarf an Elementarisierungen

In der Stochastik spielt der Kontext eine tragende Rolle; man hat es mit Modellbildung zu tun, wobei die Interpretationsmuster zwischen Realität und Modell nicht gerade direkt sind. Spiegelhalter (2014) spricht von Wahrscheinlichkeit als *virtueller* Größe und von einem metaphorischen Gebrauch von Wahrscheinlichkeit. In der beschreibenden Statistik hat der Kontext eine überragende Rolle für die Interpretation der Ergebnisse, in der beurteilenden Statistik ist die Modellsituation mit hypothetischen Modellen (Plural!) und Fehler*möglichkeiten* überfrachtet, die alle nicht vergleichbar sind. Dazu kommt, dass die verwendeten Begriffe erst innerhalb komplexerer mathematischer Zusammenhänge verstehbar werden. Die Wahrscheinlichkeitsrechnung ist zudem mit Hoffnungen und Erwartungen an die Zukunft verbunden und muss mit vielen – archetypischen – Strategien (Batanero und Borovcnik 2016) konkurrieren, die auch Zufall und Möglichkeiten für die Zukunft erklären und vorhersagen lassen; diese sträuben sich hartnäckig gegen Veränderung durch Unterweisung. Eine wirklich einfache Erfolgskontrolle fehlt mangels konsequenter Erfahrung mit dem Zufall, weil die Situationen in der Realität als Einzelfälle betrachtet werden. Es hat daher in der Stochastik schon immer einen Bedarf an Elementarisierung gegeben, nach Quellen der Einsicht, die über die Mathematik hinausgehen oder die aus anderen als mathematischen Mitteln gespeist werden.

In der Geschichte der Stochastik-Didaktik hat es schon sehr früh den Trend zur Elementarisierung hin gegeben (Borovcnik 1996). Die Explorative Datenanalyse (EDA) ist ein Paradebeispiel dafür; einerseits werden elementar verstehbare Kenngrößen verwendet, damit man die Ergebnisse im Kontext besser interpretieren kann, andererseits sucht man nach Mustern *und* Besonderheiten, um verallgemeinerungsfähige Aussagen aus Daten isolieren zu können, welche einer Bestätigung durch Methoden der beurteilenden Statistik dann gar nicht bedürfen, weil die Einsicht im Zusammenhang „synthetisch“ ist (siehe Borovcnik 1990). Das hat natürlich zu Verantwortlichkeitskonflikten geführt. Solange der Modelleur für sich Einsichten gewinnen kann, ist der Ansatz der EDA vielversprechend, sowie aber die Ergebnisse für andere verbindlich werden sollen, kommt man nicht umhin, doch wieder auf Methoden der beurteilenden Statistik zurückzugreifen, weil diese vorgeblich (modulo Modellierung, in die viel Willkür gepackt werden kann) objektive (will heißen, von der Person des Modelleurs „unabhängige“) Ergebnisse ermöglicht und – vor allem – auch eine scharfe Trennung zwischen den zu treffenden Entscheidungen erlaubt.

Für die Komplexität in der beurteilenden Statistik gibt es – von der statistischen Methodologie her – nur unzureichende Lösungen. Man weiß, dass die Überdeckungswahrscheinlichkeit eines Konfidenzintervalls nur fiktiv ist und eigentlich nur eine façon de parler darstellt. Es gibt im Normalfall keine beliebige Wiederholung einer Einzelentscheidung. Genau dasselbe trifft auf die Interpretation von α- und β-Fehler bei statistischen Tests zu. Noch dazu sind diese fiktiven Wahrscheinlichkeiten *bedingte* Wahrscheinlichkeiten. Es besteht ein weitgehender fachdidaktischer Konsens, dass Konfidenzintervalle einfacher als statistische Tests sind. Allerdings sind Fehlinterpretationen und falsche Erwartungen in ihre Wirksamkeit mindestens genauso schwerwiegend wie bei statistischen Tests. Etwa gibt die Überdeckungswahrscheinlichkeit keinen wirklichen Anhaltspunkt für die Wahrscheinlichkeit, dass die nächste Replikation des gesamten Versuchs in das berechnete Intervall fällt; diese Wahrscheinlichkeit kann sehr stark vom Konfidenzniveau abweichen (Cumming und Maillardet 2006, Gordon 2018).

Verständlich, dass es schon frühzeitig, aber nicht nur aus der didaktischen Absicht zu vereinfachen, Bestrebungen gegeben hat, die Testsituation – zumindest im ersten Zugang – zu vereinfachen (Borovcnik 1996) und die Begriffe erst später zu präzisieren. Dazu zählen auch Simulation der beteiligten Verteilungen, die Permutationstests und der Bootstrap. Die non-parametrischen Verfahren gehen auf Noether (1967) zurück; dabei braucht man keinerlei Verteilung (auch keine Normalverteilung; auch keine durch den zentralen Grenzverteilungssatz (ZGS) motivierte Annäherung durch eine Normalverteilung), außer dass es sich bei den Daten um eine Zufallsstichprobe handeln muss. Die Verfahren sind unter dem Namen Rerandomisierung bekannt: die vorhandenen Daten werden einfach permutiert. Noch weiter geht der Ansatz des Bootstraps (Efron und Tibshirani 1993), der aus vorhandenen Daten Stichproben (mit Zurücklegen) zieht und auf diese Weise eine empirische Basis für den Unterschied zweier Datensätze (die für zwei zu vergleichende Gruppen stehen) liefert. Dadurch hat man (artifiziell erzeugte) „empirische Daten“; durch Abtrennen der Extrembereiche durch Quantile erhält man die – einsichtige – Entscheidung, sprich den statistischen Test. Borovcnik (2004, 2006b) hat dies als Durchgangsstadium vorgeschlagen; nach Cobb (2007) wurden besonders in der amerikanischen Diskussion weitreichende Zugänge ausgearbeitet, die das eigentliche Ziel, die Einführung ins statistische Testen, aber geradezu vermeiden wollen (Ben-Zvi, Makar und Garfield 2018)

Jede Vereinfachung hat ihre Vorzüge und Nachteile. Für Rerandomisierung und Bootstrap, besser bekannt als *Informal Inference*, sind diese in Borovcnik (2017b) dargelegt. Ein didaktischer Nachteil ist, dass Wahrscheinlichkeit gänzlich durch relative Häufigkeiten ersetzt wird, was die genuine Vielfalt (insbesondere den Grad des Vertrauens, subjektive Wahrscheinlichkeit, Laplacesche Gleichwahrscheinlichkeit etc.) von Wahrscheinlichkeit ignoriert. Bayes-Probleme werden dadurch völlig verzerrt (siehe Borovcnik 2012, Carranza und Kuzniak 2008). Wir werden diese kritischen Erörterungen weglassen und nur auf sie verweisen; stattdessen werden wir die Vereinfachungen durchspielen.

2.2 E-Learning und Blended Learning

Neue Technologien haben die Statistik durch die Entwicklung computerintensiver Methoden radikal verändert (Borovcnik 2007a, 2007b, 2017a); sie verändern auch Lernen an sich und werden im Statistikunterricht besonders genützt. Viele Berechnungen sind nur mit Unterstützung von Software durchführbar, gar nicht zu sprechen von den Diagrammen, die dazu dienen, Ergebnisse lesbar zu machen und eine Interpretation zu unterstützen. Da der Kontext wesentlich ist, helfen auch Filme wie *Against all Odds* (Moore et al. 1995), welche authentisch zeigen, wo und wie Statistik eingesetzt wird.

Der Schritt von Software-Einsatz zur Auslagerung komplizierter Berechnungen und Diagramme zu technologisch unterstützten Lernumgebungen ist ein kleiner. Er ist dadurch angeregt, dass man in der Fachdidaktik Stochastik die Simulationsmethode aus der Fachwissenschaft als besonders geeignet ansieht, die abstrakten Wahrscheinlichkeitsmodelle durch Durchspielen – besser durch Simulation – qua ihrer Auswirkung in den relativen Häufigkeiten konkreter fassbar zu machen.

Wenn sich Lernumgebungen vom Unterricht abheben anstatt in diesen eingebunden zu sein, so ergeben sich aber Schwierigkeiten, die verhindern, dass ihr Potential ausgeschöpft wird (Wild 2007, Schuyten und Thas 2007, Nolan und Temple Lang 2007). Für die Entwicklung von Lernumgebungen wird ein systemischer Vorausblick beim Entwickler notwendig, der die Lernenden antizipieren und Weichen für den weiteren Lernvorgang bereitstellen muss. Übersieht man Fälle, die häufiger auftreten, so werden Lernende einfach steckenbleiben. Dagegen kann man in der Klasse durch ein entsprechendes Gesprächsklima Feedback bekommen, wer was und wie missverstanden oder verstanden hat (Wild 2007). Didaktisch gestützte Lernumgebungen mögen punkto Erfolg an Präsenzunterricht heranreichen; dennoch gibt es eine Präferenz für den interaktiven Face-to-Face-Kurs (Ward 2004). Allerdings darf man nicht unterschätzen, dass „Überwachung“ in Lernumgebungen, der Druck, Aufgaben regelmäßig abzugeben und diese korrigiert zurückzuerhalten, kurzfristig zu einem erhöhten Lernerfolg führen kann. So berichtet Esfandiari et al. (2010), dass Feedback über elektronisch ausgewertete Multiple-Choice- und Lückentext-Aufgaben den Erfolg der Studierenden sogar in Mega-Klassen sichert.

Der nächste Schritt ist *ein Kurs für alle*. Das umschließt einen Hypertext mit Glossar, interaktiven Diagrammen und Videos mit authentischen Anwendungen. Große Medienpreise haben die Entwicklung von E-Learning forciert (Medidaprix, GMW o.D.). Zur Statistik hat es große Projekte gegeben (etwa EMILeA o.D., Cramer et al. 2004); ein nachhaltiger Effekt ist aber ausgeblieben. Schon Härdle et al. (2006, 2007) haben die enorme Wichtigkeit von interaktiven Elementen für einen erfolgreichen Unterricht hingewiesen. So bleibt E-Learning eher ein Element, das man punktuell für Stolpersteine im begrifflichen Aufbau einsetzen kann. Eine interessante Variante hat sich in jüngerer Vergangenheit durch MOOCs (massive open online courses) ergeben, die besonders durch den Anreiz bekannter Persönlichkeiten motivierend wirken. Stellvertretend sei hier auf Kurse des Friday Institutes (o.D.) verwiesen. Als Ersatz für reguläre Programme waren sie weniger erfolgreich als es dem Einsatz der Ressourcen entspricht (The Chronicle of Higher Education o.D., Johnson et al. 2016).

Große Projekte sind weniger flexibel und können die Feedbackschleife zum Lernenden nicht so individuell gestalten, dass die Rückkoppelung den Lernerfolg langfristig absichert. Bleibt für den Unterricht an der Schule die Option Blended Learning (die Einbettung dislozierter Unterrichtsmöglichkeiten in den üblichen Präsenzunterricht), was individualisiertes Lernen wesentlich unterstützen kann. Interessant daran ist auch der Rollenwechsel für Lehrende in die Funktion eines Beraters (Borovcnik 2007b), um einige komplexere Lernschritte auszulagern.

Der normale Klassenunterricht kann u. a. durch folgende, technologisch gestützte Maßnahmen ergänzt werden: Software zur Berechnung und Erstellung von Diagrammen, dynamische Applets zur Illustration komplexer Begriffe und Methoden, Hypertext als zentraler Ausgangspunkt der Stoffvermittlung mit Erklärungen und Querverweisen auf Applets mit Illustrationen, Simulationen oder Anwendungen. In unserer langjährigen wissenschaftlichen Begleitung der Lehrveranstaltungen, insbesondere für den Bereich der Wirtschaftswissenschaften, haben wir wesentliche Kategorien identifiziert, die für den Lernerfolg entscheidend sind. Die Ansätze wurden durch Diskussion im Kollegium und mit den Studierenden verfeinert. Der Kriterienkatalog umfasst folgende Punkte (siehe Borovcnik 2017c, S. 156): #1. Entscheidung über die Software, #2. Entwicklung geeigneter Aufgaben, #3. Feedback für die Studierenden, #4. Orientierung für die verstreuten Informationen, #5. Entwickeln geeigneter Applets. Wir geben hier unsere Sicht zu den Punkten #1 und #5 wieder.

1. Die Entscheidung über die Software

> „Software ist nützlich für umständliche Berechnungen (häufig in der Statistik), für graphische Darstellungen (ein Wesen der Statistik) sowie für die Veranschaulichung komplexer Begriffe über Animation und Simulation. Extra-Aufwand zum Computing demotiviert viele Lernende […]. Wir wählen für den Einstieg Excel (XL): XL wird in vielen Kursen benötigt; XL ist in der Arbeitswelt sehr weit verbreitet; schon in der Schule sind rudimentäre Fertigkeiten darin vorhanden; XL ist einfach für die grundlegenden Dinge; und, XL ist sehr flexibel, um Effekte ‚on demand' zu zeigen, wenn eine Frage auftaucht und direkt beantwortet werden soll. Natürlich ist XL von eingeschränktem Nutzen, wenn man auf höhere Ebenen gelangt (aber nur eine Minorität wird das wirklich benötigen). Unser Fazit war: Die Entscheidung für eine Software hat einen wesentlichen Einfluss auf die Arbeitsbelastung der Studierenden. Mathematik durch Programmieraufwand zu ersetzen (wie etwa R) klingt für viele wenig überzeugend."

5. Entwickeln von geeigneten Applets

Excel muss nachgebessert werden. Dazu haben wir Templates erarbeitet, welche die Lücken schließen (Boxplot, Histogramm, Tabellierte Daten). Wir haben dynamische Animationen erstellt, um Eigenschaften von Begriffen zu klären. Was aber macht eine effektive Visualisierung aus? Recherchiert man in der Literatur, kommt man zur schnell zur Einsicht, dass Kriterien eigentlich fehlen: Sollen in Applets die Skalen automatisch angepasst werden (wie etwa in Gould o.D.), damit man das Wesentliche immer sieht, oder sollen sie starr sein, damit man Art und Größe der Veränderungen direkt sehen kann? Man gerät durch Letzteres eventuell in die Gefahr, dass der Graph aus dem Sichtbereich verschwindet, aber man erhält den Vorteil, den Einfluss eines Parameters etwa auf die Gestalt einer Wahrscheinlichkeitsverteilung direkt – ohne Reskalierung – zu sehen. Es bleibt – auch nach intensiver Diskussion – noch viel zu klären, wie man gute Applets entwickelt.

3. Dynamische Applets für den Unterricht

Wir forsten die Stochastik nach „Orten“ durch, wo durch Einsatz dynamischer Applets der Begriffserwerb nachhaltiger gestaltet wird. Wir haben vielfältige Animationen und Simulationen entwickelt, ausgetestet und verbessert. Derzeit sind sie in Lehrmaterialien eingebunden. Einige Applets sind aber so weit kommentiert, dass sie ohne weiteres in andere Kurse eingebunden werden können (Borovcnik o.D.). Sie illustrieren zentrale Fragestellungen und entsprechen didaktische Einschätzungen und Ideen, die in Batanero und Borovcnik (2016), Borovcnik und Kapadia (2011) oder Borovcnik (2011b) dargestellt werden. Als Schlüsselfaktoren der Applets nennt Borovcnik (2017c, S. 159):

> „[…] die Auswirkungen von Wahrscheinlichkeit (und damit zusammenhängenden Kennziffern) durch die Spuren der relativen Häufigkeiten zu veranschaulichen; die Relevanz von Parametern dadurch zu erforschen, dass man ihre Werte systematisch ändert und die Auswirkungen auf Verteilungen oder Ergebnisse von Stichproben untersucht; theoretische, tiefliegende mathematische Sätze (die aber erst recht wichtig sind für die Entwicklung von sekundären Intuitionen nach Fischbein 1975) in „Aktion“ zu zeigen, aber das Grenzverhalten durch ein Gedankenexperiment zu extrapolieren. Statt den Stichprobenumfang über alle Grenzen zu erhöhen (was ja gar nicht geht), soll ein stabiles Muster einer Verteilung durch Wiederholung der Simulation mit einem fixierten Stichprobenumfang (oder den Vergleich von zwei oder drei Umfängen, damit ein Trend sichtbar wird) aufgebaut werden.“

3.1 EXCEL-Kniffe

Es gibt eine Reihe von Kniffen, welche die Arbeit in Excel erleichtern, die jedoch nicht so weithin bekannt sind. Da Excel-Programmieren zu mindestens 80% aus Kopieren und Einfügen besteht, ist es wesentlich, diese Kniffe zu beherrschen. Dass man mit Excel keine ordentlichen Graphen zeichnen kann, ist ein oft vorgebrachter Kritikpunkt. Hier kann man mit einem eigenen Template Abhilfe schaffen. Mehr zu solchen Kniffen erfährt man in Borovcnik (2017c) und auf Borovcnik (o.D.)

3.2 Beschreibende Statistik

Animationen zur Darstellung der Verteilung einer Variablen

Wir geben nur einen kleinen Ausschnitt aus den Möglichkeiten (in Borovcnik o.D.), durch dynamische Änderung das Wesen von Kennziffern zu erfassen.

- Es ist didaktisch wertvoll, die Lage des Mittelwerts im Diagramm zu sehen und seine Veränderung bei Veränderung einer kleinen Gruppe von Daten zu verfolgen.
- Man kann die Änderungen des Mittelwerts bei Änderungen einiger Daten mit jenen des Medians vergleichen, um diese beiden Kenngrößen einschätzen zu lernen. Der Median bleibt gleich, es sei denn, es werden Daten über den Median hinweg „geschoben“.
- Auch kann man erkennen, wie gut oder schlecht der Mittelwert die Daten repräsentiert. Etwa markiert er nicht immer ein Zentrum der Verteilung (weil die Verteilung schief oder gar J-förmig ist) oder die Verteilung ist sehr weit gestreut um diesen Mittelwert.
- Bei Lorenz-Kurven kann man sehen, wie sich die Fläche ändert, die den Gini-Koeffizienten darstellt, wenn man die „Aufteilung“ ändert.

Templates zur beschreibenden Statistik

Wir haben Templates entwickelt, um Mängel von Excel zu beheben. Dazu zählen tabellierte Daten, Histogramme und Boxplots. Damit kann man Verteilungen zielführend untersuchen. Das gibt Rückschluss auf den Kontext und die Wirkweise der verwendeten Werkzeuge.

- Oft hat man es mit Sekundärstatistiken zu tun, wobei die Daten in Tabellen zusammengefasst sind und ein Zugriff auf die Rohdaten fehlt. Berechnungsroutinen zu tabellierten Daten fehlen in vielen Software-Paketen, auch in Excel muss man sie selbst implementieren.
- Histogramme werden in Excel wie Stabdiagramme gezeichnet (eine numerische Achse fehlt). Das stört, wenn man ungleich breite Klassen hat und oder weitere Daten oder Kennziffern (etwa die Lage des Mittelwerts) in das Histogramm einzeichnen will. Der Boxplot – eine markante Darstellung von Form und Lage einer Verteilung durch wenige Quantile – fehlt.

Animationen zur Erfassung des Zusammenhangs von zwei Variablen

- Wie man die Korrelation aus Punktwolken erkennen und wie man jene Punktwolken typisieren kann, für welche der Begriff Stärke des Zusammenhangs der beiden untersuchten Merkmale sinnvoll ist, ist mit einfachsten Methoden des Explorierens von Punktwolken möglich. Die Inspektion der Gestalt der Punktwolke sollte eine numerische Analyse immer begleiten. Wenn eine Punktwolke „homogen" erscheint, d. h., ohne Lücken ist und vertikal etwa die gleiche Breite hat, so hat der Korrelationskoeffizient durchgehend eine gute Interpretation.
- Die Punktwolke sollte in einem Quadrat gezeichnet werden, damit man die Größe des Korrelationskoeffizienten visuell beurteilen kann. Verschiedene Punktwolken (mit versteckter Skalierung) zeigen ein ganz unterschiedliches Streuverhalten (von 0,10 bis 0,90). Bei Aufdeckung der Skalierung erkennt man, dass es sich um dieselben Daten handelt, nur mit gestreckten bzw. gestauchten Skalen. Der Korrelationskoeffizient ist gegenüber linear-affinen Transformationen invariant.
- Welche Grundaufgaben den Methoden der Korrelation und Regression entsprechen, kann man durch gezielte Skizzen mit dynamischem Charakter aufzeigen. Die zentrale Problemstellung der Prognose der abhängigen Variablen weist auf die Relevanz *vertikaler* Fehler hin:

  tatsächlicher Wert minus Prognosewert (der mittels der Regressionskurve bestimmt wird).
- Wie sich die Fehler verändern, für einzelne Punkte und summarisch für den ganzen Datensatz, kann man durch Veränderung einzelner Punkte aufzeigen: Wie ändert sich dadurch die Gestalt der Punktwolke? Wie ändert sich die Regressionsfunktion (Gerade oder eine allgemeinere Funktion, das ist für Software kein Problem)? Wie ändert sich der Korrelationskoeffizient?
- Ein einfaches Applet deckt auch den grundsätzlichen Zusammenhang zwischen verschiedenen „Varianzen" und eine Additivität auf:

Totale Varianz der abhängigen Variablen *y*	=	Erklärte Varianz prognostizierter *y*-Werte	+	Nicht-erklärte Varianz der Residuen

  Diese Varianzen lassen sich einfach berechnen und man kann diese Identität feststellen; die Gleichung bleibt erhalten, auch wenn man einige Punkte im Datensatz dynamisch ändert, sodass man die Einsicht gewinnen kann, dass es sich dabei um ein mathematisches Gesetz handelt. Wenn man die Identität nachweisen will, so braucht man eine Geläufigkeit mit allgemeinen Summen und Ausdauer, um die halbe Seite mit Formeln zu einem guten Ende zu führen. Hier zeigt sich, dass mathematische Einsichten auch mit nicht-mathematischen Hilfsmitteln (Variation eines Spezialfalls) gewonnen werden können.
- Die angesprochene Additivität der Varianzen ergibt dann (auch das kann man wieder im Applet dynamisch erkennen) die folgende Deutung für das Bestimmtheitsquadrat (Quadrat des Korrelationskoeffizienten) als *anteilige Varianz*; die beiden Gleichungen zeigen, dass es dabei um ein Maß für die Güte des Regressionsmodells für die Prognose der abhängigen Variablen geht:

$$\frac{\text{Erklärte Varianz}}{\text{Totale Varianz}} = r^2, \qquad \frac{\text{Varianz der Residuen}}{\text{Totale Varianz}} = 1 - r^2.$$

- Fallstricke zu Regression und Korrelation: Simpson-Effekt, d. h., Korrelation zeigt in Subgruppen durchgehend ein zum ganzen Datensatz entgegengesetztes Vorzeichen. Korrelation ist ganz was anderes als Kausalität: eine hohe Korrelation ist ein Anlass, über kausale Zusammenhänge von der Kontextwissenschaft her nachzudenken und kein Nachweis, dass kausale Zusammenhänge bestehen. Das kann man durch paradigmatische Beispiele wie Studiendauer und Anfangsgehalt oder Zahl der Störche und Zahl der Geburten (siehe Batanero und Borovcnik 2016) zeigen. Auch hier ist begleitende Inspektion der Punktwolke instruktiv.

3.3 Wahrscheinlichkeitsrechnung

Exploration des Zufalls und Interpretation von Wahrscheinlichkeit

Im Anfangsunterricht zeigt man, dass die relativen Häufigkeiten „konvergieren“. Statt die Konvergenz der relativen Häufigkeiten zu „zeigen“, illustrieren wir, dass die relativen Häufigkeiten der Zufallszahlen 0, 1, …, 9 um die Marke 10% (die Wahrscheinlichkeit nach Laplace) schwanken, und zwar bei jeder Wiederholung völlig irregulär. Allerdings ist die Schwankungsbreite bei 1000 erzeugten Zufallszahlen weit enger als bei nur 50. In Extrapolation, quasi als Gedankenexperiment, sehen wir die Schwankungsbreite gegen Null konvergieren. Das Experiment zeigt von allem Anfang an, dass Wahrscheinlichkeiten durch relative Häufigkeiten geschätzt werden und Annahmen darüber immer gegen relative Häufigkeiten quergeprüft werden können. Für das Diagramm sei auf Borovcnik (2017c, S. 160) verwiesen.

Verständnis stochastischer Grundbegriffe

Hier sei lediglich die Additivität von Erwartungswert und Varianz angeführt: Der Erwartungswert ergibt sich bei Summenspielen additiv aus den einzelnen Spielen. Hat ein Spiel mit den Ausgängen 0 und 1 (mit Wahrscheinlichkeit p) den Erwartungswert p, so ist bei n Spielen (Anzahl der Erfolge folgen der Binomialverteilung) der Erwartungswert $n \times p$. Für die Varianz gilt Additivität allerdings nur bei Unabhängigkeit der Versuche. Damit erweist sich die Additivität für Erwartungswerte als der Linearen Algebra zugehörig, während sie für die Varianz eine genuin stochastische Eigenschaft darstellt. Wobei ein weiterer Anknüpfungspunkt zur Linearen Algebra dadurch gegeben ist, dass die Unabhängigkeit der Orthogonalität entspricht, womit die Additivität der Varianzen dem Satz von Pythagoras entspricht, wenn man die „Länge“ von Zufallsvariablen (in einem abstrakten Raum) als Wurzel aus der Varianz auffasst. Damit zeichnet man die Varianz gegenüber anderen Kennziffern für die Breite einer Verteilung aus. Wir stellen ein Applet mit Glücksrädern zur Verfügung, in dem dies exemplarisch – also nicht durch Simulation – gezeigt wird (Borovcnik 2011b).

Verteilungen und ihre Parameter explorieren

Exploration der Binomialverteilung und der geometrischen Verteilung in der Bernoulli-Kette

Zum intuitiven Grundverständnis von Wahrscheinlichkeit gehört auch, richtig einschätzen zu können, wie lange man warten muss, wenn die Wahrscheinlichkeit „bekannt“ ist. Etwa, wie lange muss man auf den ersten Sechser warten, wenn wir einen idealen Würfel werfen? Während sich kleine Änderungen der Erfolgswahrscheinlichkeit p in der Bernoulli-Kette auf die Binomialverteilung wenig auswirken, ist der Einfluss auf die Wartezeit auf den ersten Erfolg (geometrische Verteilung) drastisch. Das klärt auch den Ursprung archetypischer Vorstellungen von der kleineren Wahrscheinlichkeit des Sechsers beim Würfeln, auf den wir in frühester Kindheit „so lange“ warten mussten.

Exploration der Poisson-Verteilung

Vergrößert man den Parameter λ, so sieht man, dass sich die Verteilung symmetrisiert und der Parameter dann das Zentrum der Verteilung markiert. Mehr dazu in Borovcnik (2017c).

Exploration des Exponential-Modells

Die Exponentialverteilung hat eine J-förmige Gestalt; es hat keinen Sinn, von einem Zentrum der Verteilung zu sprechen. Aber aus der Beziehung $E(X) = 1/\lambda$ kann man den Parameter gut schätzen, weil sich die Verteilung von Mittelwerten nach dem ZGS rasch durch eine Normalverteilung approximieren lässt. Auch dies kann man durch ein dynamisches Applet illustrieren. Es reichen schon fünf exponentialverteilte Zufallsvariablen aus. Dann simuliert man entsprechend viele Stichproben und untersucht ein Histogramm der erzeugten Daten. Dies zeigt eine gute Annäherung an die Normalverteilung. Daher kann man den Parameter λ aus den Daten schätzen und aus dem resultierenden Modell ein oberes Quantil berechnen. Solche Quantile sind in der Zuverlässigkeitsanalyse viel wichtiger als der Erwartungswert, der eigentlich nur als Index zum Ansprechen des Modells dient. Solche Quantile geben Schwellenwerte an, die nur mit einer bestimmten Wahrscheinlichkeit überschritten werden. Das Komplement dieser Wahrscheinlichkeit ist dann – im Kontext von Lebensdaueranalysen – das Risiko, dass die Lebensdauer unter dem Schwellenwert bleibt. Mehr dazu in Borovcnik (2017c).

Templates

Zu allen diskreten und stetigen Verteilungen bieten wir Templates an, die den Rechenaufwand reduzieren und visualisieren, was berechnet wird. Wir zeigen, welche Bedeutung die Parameter haben, und illustrieren Berechnungen durch Flächen oder Stabsummen im Diagramm.

Stichproben – Zweck und Eigenschaften erforschen

Einfache Zufallsstichproben gegen Stichproben aus homogenen Untergruppen

Wird eine Stichprobe zufällig gezogen, so garantiert das am besten, dass sie repräsentativ wird für die Grundgesamtheit, aus der gezogen wird. Das stellt die Intuitionen auf den Kopf, insofern, als zufällig ziehen blind ziehen bedeutet, d. h., man muss von allen Eigenschaften der Objekte absehen. Muss man also alle Kenntnis über die Grundgesamtheit vergessen? Nein. Das kann man durch ein Applet demonstrieren, in welchem man aus homogenen Untergruppen zufällig zieht. Nutzt man die Kenntnis solch homogener Untergruppen aus, so kann man die Variation der Ergebnisse stark verringern und damit die Schätzungen für die Population verbessern. Mehr dazu in Borovcnik (2011b; o.D.). Als Anwendung kann man sich denken, dass aus Männern und Frauen getrennt Zufallsstichproben gezogen werden und dann die Ergebnisse aliquot auf die Grundgesamtheit hochgerechnet werden.

Link zwischen Stichprobenmittelwert und Population

Der Begriff Stichprobe kommt vom Beurteilen einer endlichen Population durch eine entsprechend gewählte Teilmenge, die Stichprobe. Sofern diese *zufällig* gezogen wird, kann man Kenngrößen der Population über Kenngrößen der Stichprobe schätzen. Es ist dabei egal, ob die Population tatsächlich endlich ist (und durch Zufallsauswahl als Zufallsvariable interpretiert wird) oder ob die Population durch eine Wahrscheinlichkeitsverteilung modelliert wird. Kern der beurteilenden Statistik ist, wie die Verteilung der Stichprobenmittelwerte (oder anderer Kenngrößen) mit der Verteilung der Population zusammenhängt. Im einfachsten Fall erhalten wir für den Mittelwert $\overline{X}$ folgende Beziehungen: $E(\overline{X}) = E(X)$ und $Var(\overline{X}) = Var(X)/n$. Hier zeigt sich, wie zentral die Additivitätseigenschaft ist. Nun mischt sich eine weitere Komplikation dazu: Von der Verteilung der Mittelwerte haben wir nur einen einzigen Wert (wir haben nur *eine* Stichprobe!). Es entspricht einem Gedankenexperiment, die Stichprobe zu wiederholen und zu untersuchen, wie ein Muster der Verteilung der Mittelwerte aussieht. Geht man von einer normalverteilten Population aus, so zeigt Borovcnik (2017c) diese Zusammenhänge. Geht man von einer allgemeinen Verteilung der Grundgesamtheit aus, so kann man ein Applet (Borovcnik 2011b) nutzen, um die Normalisierung der Verteilung der Mittelwerte zu demonstrieren (ZGS!), sodass man nur noch Erwartungswert und Varianz berücksichtigen muss.

Zentrale Sätze der Stochastik explorieren

Typen von Konvergenz und Divergenz

Es besteht – nicht nur bei Lernenden – eine Konfusion, welche aus einer Stichprobe abgeleiteten Kennziffern konvergieren und wie sich diese Konvergenz gestaltet. Wir illustrieren an dem einfachen Beispiel des wiederholten Münzwurfs die Problematik (Borovcnik 2017c, 162):

> „Wir werfen eine [Münze 20, dann 100 Mal]. Wir bestimmen dabei die folgenden Größen: *Summe von Kopf minus Zahl*, d.h., die Anzahl der Köpfe minus die Anzahl der Zahlen in der Serie (der Wert 10 bei 20 Würfen bedeutet, es wurden 15 Köpfe und 5 Zahlen geworfen); die *durchschnittliche Anzahl von Kopf minus Zahl*, d.h., wir dividieren die Differenz von Kopf und Zahl durch die Anzahl der Würfe; schließlich bestimmen wir die *standardisierte Summe von Kopf minus Zahl*. d.h., wir bestimmen zuerst Erwartungswert und Standardabweichung von Summe Kopf minus Zahl, dann ziehen wir von Summe Kopf minus Zahl den Erwartungswert ab und dividieren das Ergebnis durch die Standardabweichung.“

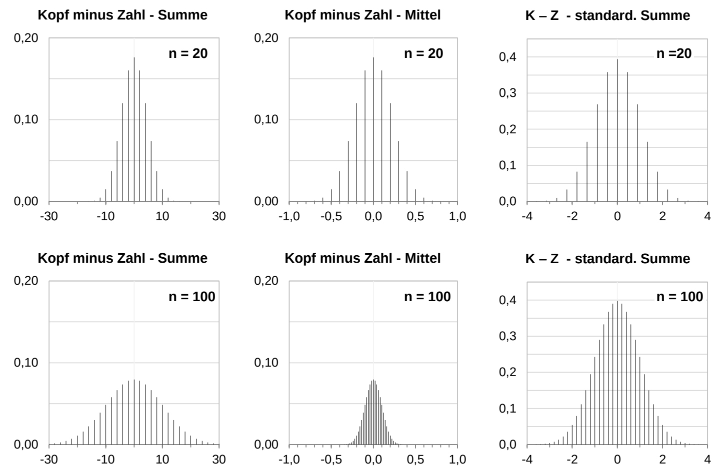


Abb.1: Divergenz der Summe – Zusammenziehen der Verteilung des Mittelwerts auf einen einzigen Punkt – Konvergenz in ein stabiles Verteilungsmuster bei den standardisierten Summen.

Das Ergebnis der (umständlichen) Berechnungen ist in Abbildung 1. Wir könnten die Muster auch durch Simulation zeigen, es wird aber durch den Störfaktor der zufälligen Variabilität schlechter sichtbar. In einem Gedankenexperiment könnte man feststellen: In der ersten Spalte (Summe) ergibt sich Divergenz, in der zweiten (Mittelwerte) zieht sich die Verteilung auf einen einzigen Punkt zusammen (Gesetz der großen Zahlen), während sich in der dritten Spalte ein stabiles Muster ergibt, das der Standard-Normalverteilung entspricht, wobei sich die Lücken zwischen den Stäben auffüllen und damit eine Approximation durch eine stetige Verteilung besser wird; Engert-Oostingh (2015) nutzt dies, um zur Integration überzuleiten. Für mathematische Details siehe Borovcnik (2011a).

Eine endliche Version des Zentralen Grenzverteilungssatzes und Approximation von Verteilungen

Bei beliebiger Verteilung eines Merkmals in der Population ergibt sich eine Normalisierung der Verteilung von Mittelwerten aus Stichproben. Hintergrund ist der Zentrale Grenzverteilungssatz (ZGS), wonach die *standardisierten* Summen der Werte aus einer Stichprobe gegen die Standard-Normalverteilung konvergieren. Für ein interessantes Experiment mit Text-Mining, das auf diese Konvergenz führt und den ZGS in überraschender Weise illustriert, sei auf Nemetz, Simon und Kusolitsch (2002) verwiesen. Wie sich diese Konvergenz auf die stabilere Gestalt der Verteilung der Mittelwerte auswirkt, kann man in den entsprechenden Applets (aus Borovcnik o.D.) nachsehen, die in Borovcnik (2017c, S. 163) ausführlich beschrieben sind. Dort wird gezeigt, wie sich die Verteilung des Mittelwerts von 10 auf 30, 50 und schließlich auf 100 Würfe eines idealen Würfels entwickelt; der Anteil von Sechsern unterliegt derselben Art von Stabilisierung.

Im Prinzip kann man die Ausgangsverteilung variieren und eine starke Schiefe oder „Zerrissenheit" einbauen. Noch immer ist das Phänomen der Normalisierung zu beobachten. Dazu ist aber eine rekursive Berechnung der Verteilung der Mittelwerte besser geeignet, weil dadurch der Störfaktor der Simulation eliminiert wird. Wie man rekursive Berechnungen ausnützt, kann man in Borovcnik und Neuwirth (2008) nachlesen. Von enormer Bedeutung ist, dass die Verteilungen nur bei standardisierten Summen *konvergieren* (ZGS). Verteilungen der Summe etwa approximiert man dann durch den ZGS asymptotisch durch eine Normalverteilung, auch wenn diese Verteilungen *divergieren*.

Stochastische Modellierung

Bayes-Probleme

Berechnungen mit der Bayes-Formel und mit verschiedensten Verteilungen kehren so oft wieder, dass es sich lohnt, ein Template zu erstellen. Dieses Template bietet neben den üblichen Berechnungen auch die visuelle Darstellung dessen, was ausgerechnet wird. Das unterstützt aktives Lernen. Es wird klar, wie sich eine Veränderung der so genannten a-priori-Wahrscheinlichkeiten und der gegebenen bedingten Wahrscheinlichkeiten auf die a-posteriori-Wahrscheinlichkeiten auswirkt. Aus dem Applet wird auch ersichtlich, dass bei sehr kleinen a-priori-Wahrscheinlichkeiten die Ergebnisse sehr sensibel auf Änderungen der Eingangsgrößen reagieren. Wie wenig die Intuition helfen kann, den Zusammenhang zwischen Eingangsgrößen und Resultaten einzuschätzen, kann man in Borovcnik (2016) im medizinischen Kontext ersehen.

Weitere Beispiele stochastischer Modellierung

Weitere Applets aus Borovcnik (o.D.) befassen sich mit folgenden Themen: Brechen von Spaghetti in drei Teile (Borovcnik und Schenk 2011), Regression zur Mitte, Risiko einer Überbelegung (etwa in einem Flugzeug), Effizienz von Spamfiltern, Telefonrechnungen, Geburten.

3.4 Beurteilende Statistik

Informelle Exploration des Zusammenhangs zwischen Stichproben und der Population

Inwieweit passt eine Stichprobe zu einem Modell der Population (egal ob die Population endlich ist oder durch eine Wahrscheinlichkeitsverteilung modelliert wird)? Man kann diese Zusammenhänge mit *Wahrscheinlichkeitsverteilungen* explorieren, indem man bei festem Stichprobenergebnis die Verteilung der Population dynamisch ändert und prüft, ob das vorgegebene Stichprobenergebnis (Faktum) in die hypothetische Verteilung „passt" oder nicht. Man kann die Verhältnisse in der Population auch *simulieren* und prüfen, wie sich bestimmte Entscheidungsregeln bewähren.

Vergleich eines beobachteten Mittelwerts aus einer Stichprobe mit einer „Population"

Es ist instruktiv, ein Diagramm der Verteilung der Population zu zeichnen und den Mittelwert einer Stichprobe darin zu markieren. Gehen wir der Einfachheit halber von einer Normalverteilung für die Population aus mit Erwartungswert $\mu = 5{,}4$ und Standardabweichung $\sigma = 1$ und einer Stichprobe vom Umfang $n = 5$, welche als Mittelwert $\bar{x} = 6$ ergeben hat. Wie gut passt dieser Mittelwert in die Verteilung der Population? Oder anders gefragt, ist ein Erwartungswert $\mu = 5{,}4$ mit der Beobachtung $\bar{x} = 6$ verträglich? In der linken Spalte von Abbildung 2 sieht man, dass das eigentlich ganz gewöhnlich ist. Nun können wir aber nicht davon ausgehen, dass in der Population tatsächlich der Erwartungswert 5,4 ist. Wir haben es hier also nur mit einem *Szenario* zu tun, wir untersuchen den Sachverhalt auf der Basis der Frage „Was wäre, wenn der Erwartungswert 5,4 zutrifft?" Wir erhalten eine Überschreitungswahrscheinlichkeit von rund 9%, will heißen, es besteht eine Wahrscheinlichkeit von fast 9%, ausgehend von unserem Szenario, solche und noch stärkere Abweichungen im Mittelwert wie in der vorliegenden Stichprobe vorzufinden.

Dieser szenario-artige Vergleich führt uns ins Zentrum der beurteilenden Statistik. Wir können aufgrund der berechneten Überschreitungswahrscheinlichkeit die Hypothese testen, ob der Wert 5,4 für die Population zutrifft und die Entscheidung an der ausgewiesenen Überschreitungswahrscheinlichkeit orientieren. Wir können aber auch dynamisch den Erwartungswert der Population verändern und prüfen, wie sich die Überschreitungswahrscheinlichkeit entwickelt. Es scheint naheliegend, einen Wert von μ = 5,0 als zu wenig plausibel zu benennen, weil die beobachtete Stichprobe nur eine Überschreitungswahrscheinlichkeit von 0,0221 für $\bar{x}$ = 6 hat. Wenn man jetzt die Werte für μ von „links nach rechts“ durchprobiert, so kann man alle Werte unterhalb von 5,1 bzw. oberhalb von 6,9 als zu wenig plausibel erachten und für die weitere Betrachtung weglassen. Genau das leistet ein Konfidenzintervall (mittlere und rechte Spalte von Abbildung 2). Die Betrachtung, ob ein fiktiver Erwartungswert für die Population mit einem beobachteten Mittelwert kompatibel ist – so wie wir das hier getan haben – führt zum Konfidenzintervall. Ändert man den Erwartungswert der Population mit einem Schieberegler dynamisch, so sieht man die Änderung der Überschreitungswahrscheinlichkeit unmittelbar. Die folgende Interpretation aus Borovcnik (2017c, S. 165) klingt jetzt sehr überzeugend:

> „Die Überlegung zeigt, dass alle Werte für den Erwartungswert der Population außerhalb dieses Intervalls das Stichprobenergebnis „kaum erklären“ können. Das Schöne ist, die Überlegungen bleiben […] zutreffend, auch wenn die Population nicht mehr als normalverteilt angenommen werden kann, weil uns dann der [ZGS] die angenäherte Normalverteilung für die Mittelwerte garantiert.“

Beurteilende Statistik vergleicht *viele* Verteilungen (die sich hier durch den Erwartungswert unterscheiden) auf Plausibilität hinsichtlich der Daten. Überprüft man einen fixen Wert μ, erhält man Tests, sucht man nach einem Filter, der nur kompatible Parameterwerte durchlässt, Konfidenzintervalle.

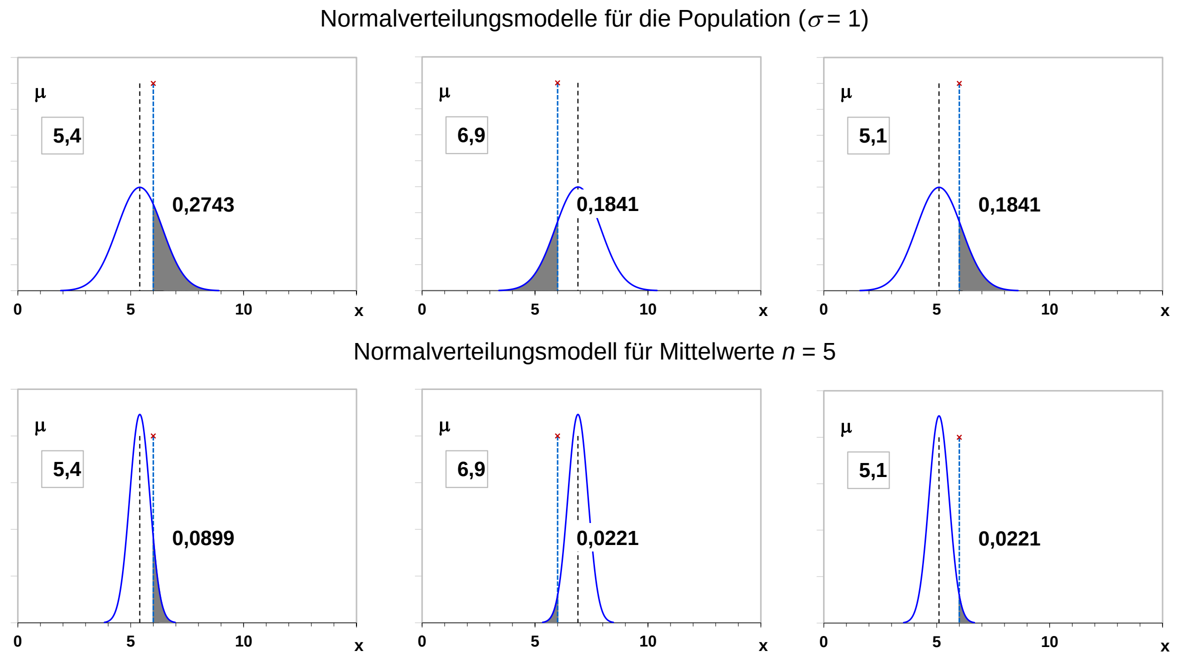


Abb. 2: Vergleich eines beobachteten Mittelwerts mit einer wechselnden Verteilung – Vorüberlegungen zu Konfidenzintervallen.

Man kann verschiedene Worte benutzen, um das Verhältnis einer speziellen Stichprobe zu unterstellten Werten für den Parameter zu beschreiben: der Wert von μ ist mit der Stichprobe verträglich, … kompatibel, … plausibel. Es wird klar, dass man Fehler machen kann, wenn man bestimmte Werte für μ ausschließt, und dass der Filter von der Größe dieser Wahrscheinlichkeit abhängt, die zum Ausschluss herangezogen wird. Analoge Überlegungen führen zu Intervallschätzungen für den normalerweise unbekannten Erfolgsparameter p beim Schätzen der Erfolgswahrscheinlichkeit in einer Bernoulli-Kette (siehe die Beschreibung des entsprechenden Applets in Borovcnik 2017c, S. 165).

Simulation formaler Konzepte und Methoden zur beurteilenden Statistik

Kontrollkarten – Interpretation von Fehlern in statistischen Entscheidungen

Ein Weg zum Testen führt über den Bedarf ein Alarmsystem einzurichten, das bei Sollbedingungen („business as usual“) nur selten Alarm auslöst, das aber bei Abweichungen, sofern diese groß genug sind, sehr häufig Alarm auslöst. Klar ist, dass wir es mit zwei Arten von Fehlern der Alarmanlage zu tun haben: Alarm, obwohl die üblichen Bedingungen zutreffen (α-Fehler); kein Auslösen des Alarms, obwohl Abweichungen bestehen (β-Fehler). In folgender Situation aus der Qualitätskontrolle kann man die Voraussetzungen des Modells gut motivieren (Borovcnik 2017c, S. 166):

> „In der laufenden Kontrolle werden stündlich 5 Werkstücke zur Prüfung entnommen. Ist ihr Mittelwert innerhalb der vorgegebenen Warngrenzen, so läuft die Produktion ungehindert weiter, fällt er außerhalb der Warngrenzen, ist aber noch innerhalb der Kontrollgrenzen, so wird nach Ursachen für die Abweichungen gesucht; fällt der Mittelwert auch außerhalb der Kontrollgrenzen, wird die Produktion abgebrochen, die Maschine wird neu eingestellt usw. Der Verlauf einer solchen Qualitätskontrolle (in Abbildung 3 von links nach rechts zeigt in grau die Abmessungen der einzelnen Werkstücke und den jeweiligen Mittelwert als Punkt. Diese Punkte müssen demnach innerhalb der gestrichelten Warngrenzen liegen, damit alles weiter geht, fallen sie außerhalb der durchgezogenen Kontrollgrenzen, so wird die Produktion unterbrochen.“

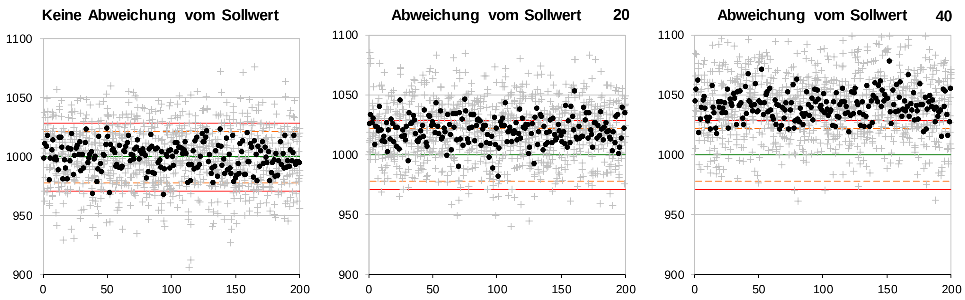


Abb. 3: Verhalten von Kontrollkarten für das Mittel von 5 Daten in 200 Stichproben – falsche und unterbliebene Alarme.

Details zum Verhalten des Alarmsystems sind in Borovcnik (2017c) nachzulesen. Der Kontext lässt viele Fragen erörtern, welche zu einem vertieften Verständnis von statistischen Tests beitragen können. Borovcnik (2017c, S. 166) resümiert:

> „Wenn man das [Szenario per Knopfdruck] erneuert, sieht man, wie stark sich die Kenngrößen ändern. Das ist [typisch]. Wenn man die besten Methoden anwendet, so hat man nicht immer denselben Erfolg, gelegentlich – aber das Risiko kann man angeben – hat man sogar Misserfolg.“

Das Applet „Werden Käfer von einer Geruchsquelle angezogen oder abgestoßen?“ behandelt eine „wissenschaftliche“ Frage auf der Basis der Binomialverteilung statt durch Simulation. Es zeigt auch die Wechselbeziehungen zwischen den (gegenläufigen) Fehlertypen (Borovcnik 2011b).

Konfidenzintervalle – Interpretation der Sicherheit

Das Konfidenzniveau von Konfidenzintervallen wird oft als Wahrscheinlichkeit, dass sich der Parameter im aufgrund der Daten der Stichprobe realisierten Intervall befindet, fehlinterpretiert. Zutreffend dagegen ist, das Konfidenzniveau als *langfristige Überdeckungsrate* zu interpretieren: wenn man die Methode (beliebig oft) wiederholt anwendet, konvergiert der Anteil der Intervalle, die den Parameter überdecken, gegen das Konfidenzniveau (Gesetz der großen Zahlen). Wie stark dieser Anteil in tatsächlichen Serien schwankt, kann man in einem entsprechenden Applet nachvollziehen, das in Borovcnik (2017c, S. 168f) beschrieben wird. Auch interessant ist der Fall, wenn die Varianz der Population aus den Daten geschätzt wird: hierbei ist die Länge der Intervalle vom Zufall abhängig.

Informal Inference: Eine Elementarisierung statistischer Inferenz

Dieser Ansatz zur statistischen Beurteilung unterstellt *keinerlei Verteilung* für die Grundgesamtheit. Wir haben mehrere Applets dazu entwickelt. Stellvertretend sei hier folgendes Beispiel gezeigt (Borovcnik 2014). Unterstellt seien je sechs Werte für die beiden Variablen x und y als Ergebnis einer Stichprobe (Abbildung 4 links). Ist die beobachtete Mittelwert-Differenz (von 33,58) signifikant? Dies kann man mit dem Zweistichproben-t-Test prüfen. Stattdessen erzeugen wir durch folgenden Vorgang die Bootstrap-Verteilung der Mittelwert-Differenz: Aus den x-Werten werden (mit Zurücklegen) sechs Werte zufällig gezogen, das wird mit den y-Werten wiederholt. Jetzt hat man eine neue „Stichprobe" mit einer neuen Mittelwert-Differenz (etwa –14,08). Danach wird derselbe Vorgang sehr oft wiederholt. Das Ergebnis von 1000 Wiederholungen zeigt die Bootstrap-Verteilung (Abbildung 4 rechts); das Ergebnis liegt knapp an der Signifikanz (5.6% der Werte sind größer oder gleich der Beobachtung).

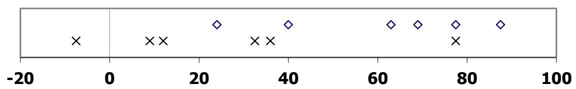
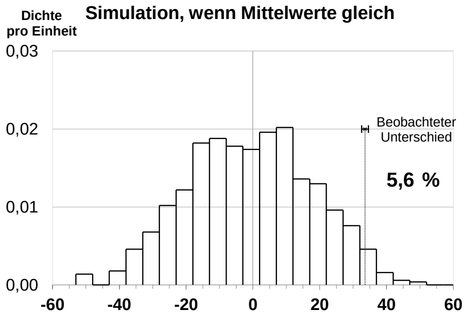


Abb. 4: Links: Versuchs- und Kontrollgruppe – Response auf der Zielvariablen (je 6 Daten).

Rechts: Bootstrap-Verteilung der Differenz und Einschätzung des beobachteten Mittelwertsunterschieds durch den p-Wert.



4. Zusammenfassung: Potential der Applets

Die Auswahl, die hier vorgestellt worden ist (oder auf die verwiesen wurde), soll illustrieren, wie dynamische Applets zum interaktiven Begriffserwerb beitragen. Für Beschreibung und Zweck, dem sie dienen und ihre Stellung in fachdidaktischen Überlegungen sei auf Borovcnik (2011b, 2017a, 2017b) verwiesen. Viele Applets sind von Borovcnik (o.D.) abrufbar. Wir fassen noch Aspekte stochastischer Begriffe zusammen, die uns wesentlich sind und die durch Applets sichtbar werden.

4.1 Stochastische Schlüsselbegriffe, die durch die Applets illustriert werden

Der fachdidaktische Mehrwert von Applets richtet sich danach, formale Mathematik zu umgehen und doch mathematische Einsichten zu ermöglichen. Borovcnik (2017c, S. 168) fasst das so zusammen:

> „An Einflussgrößen, welche das Potential der Applets ausschöpfen, sind uns […] erwähnenswert:
> - die Auswirkung von Wahrscheinlichkeit (und verwandter Größen) durch relative Häufigkeiten zu zeigen;
> - die Relevanz einzelner Parameter durch Änderung derselben und Nachverfolgen der Auswirkungen auf andere Begriffe (etwa das Aussehen der Verteilung) zu demonstrieren;
> - mathematische Sätze nachspielen und damit ihre Aussage bzw. Wirkung verstehen;
> - Grenzverhalten durch Gedankenexperimente zu extrapolieren; statt den Stichprobenumfang ad infinitum zu erhöhen, wird bei wiederholter Simulation bei *festem* Stichprobenumfang eine *Gestalt* der Verteilung erkennbar."

Natürlich hängen diese Faktoren mit einer bestimmten Sicht von Wahrscheinlichkeit und Statistik zusammen, die der Autor in der Monographie zur *Stochastik im Wechselspiel zwischen Intuition und Mathematik* erstmals zusammenhängend dargestellt (Borovcnik 1992) und später (Borovcnik 2006a) verfeinert hat. Dazu kommt, dass die Deutung von Wahrscheinlichkeit als relative Häufigkeiten ohne Grenzwertüberlegungen aufgebaut wird. Grenzwerte entsprechen immer einer genuin theoretischen Sicht der Dinge und können realiter nie eingelöst werden, weder in einer Bestätigung der Konvergenz noch in der Erfassung der Nähe zum Grenzwert und schon gar nicht, wo der Grenzwert liegen wird.

Zu solchen Überlegungen sind die Folgen der relativen Häufigkeiten absolut ungeeignet aufgrund der Unabhängigkeit, will heißen, aufgrund fehlender Bildungsgesetze wie sie in der Analysis auftreten.

> Die vorgestellten Gedankenexperimente deuten in Richtung des Erkennens stabiler Muster bei der Erhöhung der Länge der Serie (des Stichprobenumfangs) in wenigen exemplarischen Fällen. Daraus soll ein „Trend“ erkannt und gedanklich fortgesetzt werden. Dieser Ansatz ist besonders mit Applets schön zu verfolgen und zeigt eine ganz wesentliche Eigenschaft von Wahrscheinlichkeit und statistischen Methoden: Mit größeren Stichproben, sofern diese wirklich auf dem Zufall basieren, wird die statistische Fluktuation des Ergebnisses immer weniger wichtig; das heißt, die Ergebnisse einer Stichprobe spiegeln die Verhältnisse in der Population sehr genau.“ (Borovcnik 2017c, S. 170)

Das bedeutet, schon bei der Klärung der Deutung von Wahrscheinlichkeit als relative Häufigkeit wird ein Gedankenexperiment vorbereitet, das dann – analog – auf den Schluss von der Stichprobe auf eine Grundgesamtheit in der statischen Inferenz fortgesetzt werden kann. So wie man Wahrscheinlichkeiten aus endlichen (zufälligen) Stichproben immer besser schätzen kann, wenn die Stichproben größer werden, so steht es auch mit jedem Parameter einer Population, die qua zufälliger Stichprobe zu einer Wahrscheinlichkeitsverteilung wird: Größere Stichproben lassen diese Parameter besser schätzen. Dabei vermeiden wir eine Reduktion von Wahrscheinlichkeit auf die Deutung als relative Häufigkeit. Der hypothetische Charakter von Wahrscheinlichkeit kommt insbesondere in den stochastischen Modellierungen sowie im Explorieren von Verteilungen im Sinne der informellen Vorbereitung der Methoden der statistischen Inferenz zum Tragen. Da wird immer im Sinne von „was wäre denn, wenn diese Verteilung zuträfe?“ (im Konjunktiv!) gesprochen. Wir verwenden Verteilungen immer unter der Prämisse einer Möglichkeit und vergleichen viele Verteilungen im Hinblick auf reale Konsequenzen miteinander. So erhalten wir auch einen Einblick in die diversen Folgen von Entscheidungen.

4.2 Applets als Bereicherung der Ausbildung in Stochastik in der Schule

Der spielerische Umgang mit den Applets ist uns wesentlich. Das unterstützt Lernen als interaktiven Prozess. Die Konsequenzen der Modelle werden sichtbar und damit werden die abstrakten stochastischen Begriffe mit Leben erfüllt. Explorieren der Modelle soll dazu beitragen, den Sachverhalt in seiner Aussage und Relevanz besser zu verstehen. Ziel ist es zu illustrieren, dass man Mathematik nicht nur mit mathematischen Hilfsmitteln erschließen kann. Mit Erkenntnissen auf der Meta-Ebene werden mathematische Einsichten gefördert. Wie die Applets in eine einführende Lehrveranstaltung in Statistik für Wirtschaftswissenschaften an der Universität eingebettet werden, kann man aus dem Projekt *New Technologies in Statistics Education*, das auf ResearchGate dargestellt wird, nachvollziehen. Viele Applets sind zugänglich (Borovcnik o.D.). Sie werden laufend durch Rückfragen und Rückmeldungen adaptiert. Die Hoffnung besteht, dass damit ihre Praxistauglichkeit verbessert wird.

Der Autor dankt Herrn Franz Schoberleitner für die kritische Durchsicht des Beitrags, welche die Lesbarkeit der Ideen erheblich verbessert hat.

Literatur

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

Ben-Zvi, D .; Makar, K.; Garfield J. (Eds.) (2018): *International handbook of research in statistics education*. Cham: Springer.

Borovcnik, M. (o.D.): *Excel-Files für den Unterricht. Sammlung virtueller statistischer Experimente*. Online: wwwg.uni-klu.ac.at/stochastik.schule/Boro/index_inhalt (Zugriff: 20.12.2018).

Borovcnik, M. (1990): Explorative Datenanalyse – Techniken und Leitideen. *Didaktik der Mathematik, 18*, 61–80.

Borovcnik, M. (1996): Trends und Perspektiven in der Stochastik-Didaktik. In: G. Kadunz; H. Kautschitsch; G. Ossimitz; E. Schneider (Hrsg.): *Trends und Perspektiven*. Wien: Hölder-Pichler-Tempsky, 39–60.

Borovcnik, M. (1992): *Stochastik im Wechselspiel von Intuitionen und Mathematik*. Mannheim: Bibliographisches Institut.

Borovcnik, M. (2004): Resampling mit Excel. *Stochastik in der Schule, 24*(3), 22–27. Bearbeitung von Christie, D. (2004): Resampling with Excel. (*Teaching Statistics, 26*, 9–14).

Borovcnik, M. (2006a): Probabilistic and statistical thinking. *Proceedings of CERME 4*. Barcelona: ERME, 484–506.

Borovcnik, M. (2006b): On outliers, statistical risks, and statistical inference. In: R. Biehler, D. Pratt (Hrsg.): Working Group 5 „Stochastic Thinking". Paper presented at CERME 5, Larnaca. Online: wwwg.uni-klu.ac.at/stochastik.schule/Boro/index_inhalt (Zugriff: 20.12.2018).

Borovcnik, M. (2007a): The influence of software support on stochastics courses. In: M.I. Gomes; J.A. Pinto Martins; J.A. Silva (Hrsg.): *Bulletin 56. Session of the International Statistical Institute (ISI)*. Voorburg: International Statistical Institute, 4602–4605.

Borovcnik, M. (2007b): New Technologies revolutionize the applications of statistics and its teaching. In: M.I. Gomes; J.A. Pinto Martins; J.A. Silva (Hrsg.): *Bulletin 56. Session of the International Statistical Institute*. Voorburg: International Statistical Institute, 823–826.

Borovcnik, M.; Neuwirth, E. (2008): Rekursive Zugänge zu Wahrscheinlichkeitsproblemen und ihr Potential zur Modellbildung. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft, 41*, 1–20.

Borovcnik, M. (2009): Aufgaben in der Stochastik – Chancen jenseits von Motivation. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft, 42*, 20–42.

Borovcnik, M. (2011a): *Stochastik 1 für Mathematik- und Informatik-Studierende*. Klagenfurt: Alpen-Adria-Universität.

Borovcnik, M. (2011b): Key properties and central theorems in probability and statistics – Corroborated by simulations and animations. *Selçuk Journal of Applied Mathematics. Special Issue of Statistics*, 3–19.

Borovcnik, M. (2012): Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didactica de la Matemática, 2*, 5–27.

Borovcnik, M. (2014): Vom Nutzen artifizieller Daten. In U. Sproesser; S. Wessolowski; C. Wörn (Eds.), Daten, Zufall und der Rest der Welt. Berlin: Springer, 27–43.

Borovcnik, M. (2016): „To screen or not to screen" – Dialoge zur medizinischen Diagnose. *Mathematik lehren, 194*, 22–28.

Borovcnik, M. (2017a): New technologies, software, and e-learning – Enriching courses in introductory statistics and probability. *Uygulamalı Sosyal Bilimler Dergisi (Journal for Applied Social Sciences), 1*(2), 62–78.

Borovcnik, M. (2017b): Informal inference – Some thoughts to reconsider. *Proceedings of the 61st World Statistics Congress*. The Hague: International Statistical Institute (ISI).

Borovcnik, M. (2017c): Applets als Bereicherung der Ausbildung in Stochastik in der Schule. *Mathematik im Unterricht, Ausgabe 8*, 155–171.

Borovcnik, M.; Kapadia, R. (2011): Modelling in probability and statistics – Key ideas and innovative examples. In: J. Maaß; J. O'Donoghue (Hrsg.): *Real-world problems for secondary school students – Case studies.* Rotterdam: Sense Publishers, 1–44.

Borovcnik, M.; Schenk, M. (2011): Simulationen im Stochastik-Unterricht. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft, 44*, 1–16.

Carranza, P.; Kuzniak, A. (2008): Duality of probability and statistics teaching in French education. In C. Batanero; G. Burrill; C. Reading; A. Rossman (Hrsg.): Proceedings of the Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Granada: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf (Zugriff: 20.12.2018).

Cobb, G.W. (2007): The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1–15. Online: escholarship.org/uc/item/6hb3k0nz (Zugriff: 20.12.2018).

Cramer, K.; Kamps, U.; Zuckschwerdt, C. (2004): St Apps and EMILeA-Stat. Interactive visualizations in descriptive statistics. In: J. Antoch (Hrsg.): *Proceedings in computational statistics*. Heidelberg: Physica, 101–112.

Cumming, G.; Maillardet, R. (2006): Confidence intervals and replication: Where will the next mean fall? *Psychological Methods, 11*, 217–227.

Efron, B.; Tibshirani, R.G. (1993): *An introduction to the Bootstrap*. New York: Chapman and Hall.

EMILeA (o.D.): *Eine multimediale, internetbasierte, interaktive Lehr- und Lernumgebung in der angewandten Statistik.* Online: www.emilea.de/ (Zugriff: 20.12.2018).

Engert-Oostingh, I. (2015): *Ein genetisch orientierter Lehrgang zur Wahrscheinlichkeitsrechnung*. Salzburg: Universität (Unveröffentlichte Dissertation).

Esfandiari, M.; Nguyen, H.; Yaglovskaya, Y.; Gould, R. (2010): Enhancing statistical literacy through short open-ended questions that involve context, data, and upper level thinking. In: C. Reading (Hrsg.): *Data and context in statistical education: Towards an evidence-based society*. Voorburg: International Statistical Institute.

Fischbein, E. (1975): The intuitive sources of probabilistic thinking in children. Dordrecht: Reidel.

Friday Institute (o.D.): *MOOC-Ed: Massive open online courses for educators*. NC State University. Online: place.fi.ncsu.edu/ (Zugriff: 20.12.2018).

GMW (o.D.): Gesellschaft für Medien in der Wissenschaft, Medidaprix.

Gomes, M. I.; Pinto Martins, J. A.; Silva, J. A. (2007) (Hrsg.): *Bulletin 56. Session of the International Statistical Institute (ISI).* Voorburg: International Statistical Institute.

Gordon, I.R. (2018): Confidence intervals and replication: The reality. In M.A. Sorto; A. White; L. Guyot (Hrsg.): *Looking back, looking forward.* Auckland: International Association for Statistics Education. Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018 (Zugriff: 20.12.2018).

Gould, R. (o.D.): Visualisation of the normal distribution. *Statistics Online Computational Resource (SOCR).* Online: socr.stat.ucla.edu/htmls/SOCR_Distributions.html (Zugriff: 20.12.2018).

Härdle, W.; Klinke, S.; Ziegenhagen, U. (2006): *E-learning statistics – A selective review.* SFB 649 Discussion Paper 2006-024. Berlin: Humboldt University.

Härdle, W.; Klinke, S.; Ziegenhagen, U. (2007): On the utility of e-learning in statistics. *International Statistical Review, 75*(3), 355–364.

Johnson, L.; Adams Becker, S.; Cummins, M.; Estrada, V.; Freeman, A.; Hall, C. (2016): *NMC Horizon Report: 2016 Higher Education Edition.* Austin, TX: The New Media Consortium. Online: cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf (Zugriff: 20.12.2018).

Moore, D.; Braun, S.; Blatt, J.; Girshman, P.; Singer, J.; Rubin, J.; Stewart, S.; Huntley D.; Buderi, R. (1995): *Against all odds: Inside statistics.* Online: www.learner.org/courses/againstallodds/unitpages/index.html (Zugriff: 20.12.2018).

Nemetz, T.; Simon, J.; Kusolitsch, N. (2002): Überzeugen statt Beweisen – der zentrale Grenzverteilungssatz im Gymnasialunterricht. *Stochastik in der Schule 22*(3), 4–7.

Noether, G.E. (1967): *Elements of nonparametric statistics*. New York: J. Wiley.

Nolan, D.; Temple Lang, D. (2007): Dynamic, interactive documents for teaching statistical practice. *International Statistical Review, 75*(3), 295–321.

Schuyten, G.; Thas, O. (2007): Statistical thinking in computer-based learning environments. *International Statistical Review, 75*(3), 365–371.

Spiegelhalter, D. (2014, April). What can education learn from real-world communication of risk and uncertainty? *Invited lecture at the Eight British Congress on Mathematical Education*, Nottingham.

The Chronicle of Higher Education (o.D.): *The minds behind the MOOCs*. Online: chronicle.com/article/The-Professors-Behind-the-MOOC/137905/?cid=at&utm_source=at&utm_medium=en#id=overview (Zugriff: 20.12.2018).

Ward, B. (2004): The best of both worlds: A hybrid statistics course. *Journal of Statistics Education, 12*(3), 74–79.

Wild, C. (2007): Virtual environments and the acceleration of experiential learning. *International Statistical Review, 75*(3), 322–335.

Verfasser

Manfred Borovcnik, Alpen-Adria-Universität Klagenfurt, Institut für Statistik, Universitätsstraße 65, 9020 Klagenfurt

manfred.borovcnik@uni-klu.ac.at