

Ein Witz! Kritik und Gegenkritik zur Nowitzki-Aufgabe im NRW Zentralabitur 2008

RAPHAEL DIEPGEN, BOCHUM

Zusammenfassung: Die stochastikdidaktische Kritik an der fachlichen Kritik der so genannten Nowitzki-Aufgabe im Zentralabitur Nordrhein-Westfalen 2008 offenbart dieselben „witzigen“ Wirrungen wie die Aufgabe selbst.

Und noch einmal: Die Aufgabe

Im NRW-Zentralabitur 2008 wurde folgende (hier nur in einem kleinen Auszug wiedergegebene) Stochastikaufgabe gestellt:

„Der deutsche Basketball-Profi Dirk Nowitzki spielt in der amerikanischen Profiliga NBA beim Club Dallas Mavericks. In der Saison 2006/2007 erzielte er bei Freiwürfen eine Trefferquote von 90,4 %.

- a) Berechnen Sie die Wahrscheinlichkeit, dass er
- (1) genau 8 Treffer bei 10 Versuchen erzielt,
 - (2) höchstens 8 Treffer bei 10 Versuchen erzielt,
 - (3) höchstens vier Mal nacheinander bei Freiwürfen erfolgreich ist.
- b) Bei Heimspielen hatte er eine Freiwurfbilanz von 267 Treffern bei 288 Versuchen, bei Auswärtsspielen lag die Quote bei 231:263. Ein Sportreporter berichtet, dass Dirk Nowitzki auswärts eine deutlich schwächere Freiwurfquote habe.

Untersuchen Sie auf einem Signifikanzniveau von 5 %, ob die Trefferanzahl bei Auswärtsspielen

- (1) signifikant unter dem Erwartungswert für Heim- und Auswärtsspiele liegt,
- (2) signifikant unter dem Erwartungswert für Heimspiele liegt.

(Hinweis: Für eine binomialverteilte Zufallsgröße X mit Standardabweichung $\sigma > 3$ gilt näherungsweise $P(X \geq \mu - 1,64\sigma) \approx 0,95$.)

c) ... d) ... e) ...“

Diese – inzwischen als „Nowitzki-Aufgabe“ berühmt-berüchtigte – Abituraufgabe wurde wegen der (zumindest) unklaren Aufgabenstellung in a) (3) zum Thema in der skandalisierenden Tagespresse, aber auch in Gänze, insbesondere in der zitierten Teilauf-

gabe b), Gegenstand von nunmehr drei Beiträgen in dieser Zeitschrift.

Meine Kritik

Sofort nach Bekanntwerden der Aufgabe veröffentlichte ich eine ausführliche Kritik (Diepgen, 2008). Neben der ins Auge springenden inhaltlichen Irrelevanz der Problemstellung und der bereits öffentlich skandalisierten Unklarheit in a) (3) kritisierte ich vor allem folgende zentrale Punkte:

1. Die Aufgabe lasse sich kaum inhaltlich so interpretieren, dass die – von den Aufgabenautoren als selbstverständlich unterstellte – Binomialverteilung plausibel wäre (Stichworte: Konstanz der Trefferwahrscheinlichkeit, stochastische Unabhängigkeit). Insbesondere sei kaum zu verstehen, was man hier eigentlich unter dem Parameter dieser Binomialverteilung verstehen solle und könne, also unter der „Trefferwahrscheinlichkeit“ von Dirk Nowitzki.
2. Es sei völlig unklar, ob die in der Freiwurfbilanz angegebenen Trefferquoten als (feste) Parameter der unterstellten Binomialverteilungen, also als Wahrscheinlichkeiten, zu interpretieren seien, oder aber als (zufällige) Schätzer für diese Parameter, also als relative Häufigkeiten. In ein und demselben Kontext figurierten dieselben Trefferquoten sowohl als Parameter als auch als Schätzer.
3. Diese Verwirrung resultiere schließlich in der Absurdität, dass als „richtige“ Lösung zu Aufgabe b) ein und dieselbe Nullhypothese aufgrund derselben empirischen Daten auf ein und demselben Signifikanzniveau in (1) beibehalten, in (2) aber verworfen werde.

Des Weiteren kritisierte ich für die Aufgabenteile c) und d) die mangelnde Konsequenz in der nur unklaren und halbherzigen Orientierung an den – nicht kompatiblen – Grundideen der miteinander konkurrierenden inferenzstatistischen Schulen von Fisher (Stichwort: Unplausibilität einer Nullhypothese durch Daten mit geringer Überschreitungswahrscheinlichkeit) versus Neyman und Pearson (Stichwort: Verfahren zur datengestützten Entscheidung zwischen zwei konkurrierenden Hypothesen mit limitierten bzw. kontrollierten Wahrscheinlichkeiten für Fehlentscheidungen

der ersten bzw. zweiten Art) versus Bayes (Stichwort: Modifikation von Hypothesenwahrscheinlichkeiten aufgrund von Daten).

Ausführlich thematisierte ich auch die grundsätzliche Problematik, dass sich die Erreichung wesentlicher Lernziele des Stochastikunterrichtes – etwa die „vernünftige“ Orientierung in komplexen Situationen der Ungewissheit – kaum in der simplen Form von landeseinheitlichen Abituraufgaben überprüfen lasse und dass von daher durch das Zentralabitur eine Trivialisierung anspruchsvoller Lernziele des Stochastikunterrichtes zu befürchten sei.

Die Kritik von Davies

So als habe es meinen Beitrag gar nicht gegeben, erschien im Folgejahr ein Artikel von Davies (2009), also eines universitären Statistikprofessors. Im Wesentlichen kritisierte er – garniert durch viele professorale Belehrungen über die Basics und Begrifflichkeiten der an den Universitäten betriebenen mathematischen Statistik – nur die verwirrende Doppelbedeutung der Trefferquoten gleichzeitig als Parameter („wahre Werte“) und Schätzer, also halt nur das, was ein einziger Kritikpunkt (2) von mir war. Die anderen von mir skizzierten Probleme gerieten Davies „natürlich“ nicht in den Blick, weil er ausschließlich aus einer innermathematischen Perspektive argumentierte: Aus dieser engen Perspektive spielt die stochastische Situation, die eigentlich erst modelliert werden soll, kaum eine Rolle; aus dieser Perspektive geht es vielmehr nur noch um die logischen Folgerungen aus einer bereits vorgegebenen Modellierung. (Und da endet man dann als Statistikprofessor halt in einem empörten, mathematisch geschulten „Kannitverstan“, wenn einem dieselben Werte als „wahre“ Werte und zugleich als Schätzwerte präsentiert werden.)

Diese Perspektive ist ärgerlich, denn im allgemeinbildenden Stochastikunterricht geht es – anders als möglicherweise in universitären Stochastikvorlesungen für Mathematikstudenten – wesentlich nicht (nur) um innermathematische Argumentationen (und tunlichst dem Computer überlassene Berechnungen) innerhalb eines vorgegebenen mathematischen Modells, sondern zuvörderst um die Modellierung selbst, also die Entwicklung und kritische Reflexion eines mathematischen Modells der jeweiligen Realität, in der Stochastik typischer Weise eines mathematischen Modells einer bestimmten stochastischen Situation der Ungewissheit. Zitat Davies (S. 2): „Bei einer Textaufgabe (*sic!*) muss die Modellierung entweder angegeben werden, oder sie muss aus der Beschreibung der Situation eindeutig (*sic!*) hervorgehen.

Wenn die Modellierung feststeht, wird die Aufgabe innerhalb des Modells weiter bearbeitet. D. h. es wird angenommen, dass das Modell stimmt. Es ist nicht Aufgabe des Abiturienten, die Angemessenheit der Modellierung zu hinterfragen.“ Sic! Hat Davies von der stochastikdidaktischen Diskussion der letzten Jahrzehnte eigentlich irgendeine Notiz genommen?

(Übrigens: Davies scheint nicht nur die stochastikdidaktische Diskussion kaum zu kennen, sondern auch die Grundlagendiskussion zur Statistik: Anders ist seine apodiktische Behauptung kaum zu verstehen, „in der Statistik“ gebe man für einen unbekannt Parameter p „ein so genanntes Konfidenzintervall von plausiblen Werten p “ an, und deshalb ergebe sich – aufgrund des 95 %-Konfidenzintervalles $[0.8792, 0.9284]$ um die berichtete Trefferquote $498/551 = 0.904$ – als „einzig plausible Lösung“ für die in a) (1) erfragte Wahrscheinlichkeit „die Behauptung, dass die gesuchte Wahrscheinlichkeit zwischen 0.2345 ($p = 0.8792$) und 0.1273 ($p = 0.9284$) liegt.“ (S. 4) Es ist schon bemerkenswert, wie hier Davies eine in der Grundlagendiskussion umstrittene Vorgehensweise – Konfidenzintervall – zur allgemeinverbindlichen Norm erhebt, und dabei auch noch das Niveau 95 %.)

Die Kritik der Kritik durch Borovcnik

Da konnte eine Reaktion seitens der Stochastikdidaktik nicht lange auf sich warten lassen: Borovcnik (2009) weist aus dieser Perspektive die Kritik von Davies zurück, mündend in dem vernichtenden Fazit: „Die Kritik hält einer Kritik fast weniger Stand als die kritisierte Aufgabe selbst.“ (S. 16)

(Meinen Beitrag listet Borovcnik übrigens seltsamer Weise in seinem Literaturverzeichnis auf, ohne ihn im Text irgendwo zu erwähnen.)

Mit Bezug zur Aufgabe a) moniert Borovcnik die – von Davies nicht, von mir aber sehr wohl kritisierte – Selbstverständlichkeit der von den Aufgabenautoren unterstellten Modellierung mit der wenig angemessenen Binomialverteilung. (Warum diese Unangemessenheit im Wesentlichen nur Aufgabe a) betreffe, kaum mehr aber Aufgabe b), habe ich freilich nicht verstanden – wie auch so manch andere „dunkle“ Bemerkung von Borovcnik.)

Die Aufgabe b) aber, die von Davies ebenso wie von mir (schon) wegen der offensichtlichen Vermischung von Schätzern und Parametern als unsinnig und unlösbar kritisiert wurde, wird von Borovcnik verblüffender Weise als durchaus sinnvoll akzeptiert und dann im Sinne der Aufgabenautoren „gelöst“.

Dazu behauptet Borovcnik (S. 16) allen Ernstes: „Die Frage in b) kann man ... so umformulieren: Vorausgesetzt man kennt diese Wahrscheinlichkeit,“ – gemeint ist die „Wahrscheinlichkeit, einen Freiwurf zu realisieren“ – „ist es dann möglich, die Auswärtsspiele als

- unabhängige Wiederholung desselben Bernoulli-Experiments
- mit eben dieser Wahrscheinlichkeit (für alle Spiele bzw. noch besser für Heimspiele)

zu modellieren? Weichen die beobachteten Trefferzahlen (oder Trefferquoten, das ist gleichwertig) signifikant davon ab oder nicht? – das ist die Frage. Dabei taucht die Erfolgswahrscheinlichkeit für die Bezugsverteilung (am besten die Heimspiele) als feste Größe auf, für die man nach der Saison einen bekannten Wert hat. Die beobachtete Erfolgsrate oder Erfolgsanzahl für die Auswärtsspiele ist eine empirische Größe, deren beobachteter Wert mit dieser Bezugsverteilung verglichen wird.“

So lässt sich Aufgabe b) tatsächlich aber eben nicht verstehen! Für die in Aufgabe b) skizzierte stochastische Situation gibt es nämlich nicht den geringsten vernünftigen oder auch nur nachvollziehbaren Grund, die Trefferquote 267/288 bei Heimspielen als „feste Größe“ oder „bekannten Wert“ – also als Parameter oder Wahrscheinlichkeit – zu interpretieren, die *völlig gleich geartete* Trefferquote 231/263 bei Auswärtsspielen aber als „empirische“, also zufällige Größe oder Schätzer. Natürlich *kann* man „rein theoretisch“ eine solche Fiktion gedanklich durchspielen: Nur gibt die in Aufgabe b) skizzierte stochastische Situation nicht den geringsten vernünftigen Anlass für eine solche Fiktion. Nein, diese Situation lässt es vernünftigerweise nur zu, die Trefferquoten – auswärts wie zuhause – entweder *beide* als feste Parameter zu interpretieren, oder aber *beide* als empirische Zufallsgrößen (Schätzer).

Im ersten Fall ist die Frage nach einem „signifikanten“ Unterschied zwischen den Trefferquoten – sprich: festen Parametern – völlig sinnlos: Parameter unterscheiden sich oder eben nicht, aber niemals „signifikant“; sind die Parameter bekannt, ist es schließlich schon vor jeder denkbaren Datenerhebung sicher, ob die Nullhypothese gilt oder nicht.

Im zweiten Fall mag diese Frage nach dem „signifikanten“ Unterschied zwischen den Trefferquoten zwar Sinn haben; sie lässt sich dann aber von den Abiturienten nicht beantworten, weil diese halt nur den Binomialtest kennen, nicht aber den Chi-Quadrat-Test, den exakten Test von Fisher oder gar den

Welch-Test. Im Übrigen wird diese zweite Interpretation durch das Wort „Erwartungswert“ im Aufgabentext ausgeschlossen; dieser ist schließlich per allgemein anerkannter Definition das n -fache des festen Wahrscheinlichkeitsparameters, nicht etwa des zufälligen Schätzers dafür.

Dass gerade von einem Stochastikdidaktiker versucht wird (und dies auch noch unter dem Titel richtig verstandener Anwendungsorientierung), die Aufgabe b) – völlig losgelöst von der dort skizzierten stochastischen Situation und ohne Angabe irgendwelcher sich daraus ergebenden vernünftigen oder auch nur nachvollziehbaren Gründe – als legitime Aufforderung für eine fiktive formale Spielerei zu reinterpretieren, finde ich befremdlich. Wenn denn das Lehrziel des Stochastikunterrichtes die vernünftige Orientierung in stochastischen Situationen ist, verbieten sich doch völlig ungläubwürdige Textaufgaben, die – bar jeder praktischen „Vernunft“ – nur der „Einkleidung“ irgendwelcher formaler mathematischer Inhalte dienen.

Besonders ärgerlich ist dieser unhaltbare Legitimationsversuch einer unhaltbaren Abituraufgabe auch, weil er ohnehin nur für die zweite Teilaufgabe b) (2) greifen könnte, nicht aber für die erste Teilaufgabe b) (1) – wie auch Borovcnik selbst mit seinem vagen Einschub „am besten die Heimspiele“ zu spüren scheint. Denn mit Bezug zur ersten Teilaufgabe b) (1) müsste Borovcnik die Auswärtstrefferquote 231/263 als Zufallsgröße und die Gesamtrefferquote $(231 + 267)/(263 + 288)$ als festen Parameter interpretieren – obwohl sich diese vermeintlich feste Gesamtrefferquote als gewichtete Summe aus (als fest interpretierter) Heimtrefferquote und – gerade als zufällig vorausgesetzter – Auswärtstrefferquote zusammensetzt: Wie kann es (in der Nichtbayesschen Statistik) einen festen Parameter geben, der auch – gleichsam „zur Hälfte“ – zufällig ist? Völlig absurd!

Diese von Borovcnik augenscheinlich nicht bemerkte Absurdität gipfelt schließlich darin, dass er – ganz im Sinne der Aufgabenautoren – für die erste Teilaufgabe b) (1) keinen signifikanten Unterschied ermittelt („Es kann statistisch (auf dem Signifikanzniveau 5 %) nicht nachgewiesen werden, dass Nowitzki in Auswärtsspielen schlechter ist als in der gesamten Saison“), für die zweite Teilaufgabe b) (2) völlig unbekümmert davon aber einen signifikanten Unterschied vermeldet („Auswärtsspiele unterscheiden sich ... hochsignifikant von Heimspielen“). Hallo? Wir sollen also nach der Logik des Hypothesentestens aufgrund der in Aufgabe b) skizzierten Datenlage davon ausgehen, dass sich einerseits die Trefferwahr-

scheinlichkeit Nowitzkis zwischen Auswärts- und Heimspielen unterscheidet, dass sich andererseits die Trefferwahrscheinlichkeit Nowitzkis zwischen Auswärts- und allen Spielen aber *nicht* unterscheidet, obwohl sich diese beiden Aussagen gegenseitig logisch ausschließen? Das kann – im wahrsten Sinne des Wortes – nicht wahr sein! Borovcnik realisiert – ebenso wie die Aufgabenautoren – offensichtlich gar nicht, dass die in b) (1) und b) (2) getesteten Nullhypothesen äquivalent sind, dass also von ihm in b) (2) genau die Nullhypothese verworfen wird, die in b) (1) – aufgrund derselben Daten und desselben Signifikanzniveaus – nicht verworfen wird.

Formal:

$$p_A = \frac{n_H}{n_H + n_A} p_H + \frac{n_A}{n_H + n_A} p_A \quad (1)$$

$$\Leftrightarrow p_A = p_H \quad (2)$$

mit den durch H („Heim“) und A („Auswärts“) indizierten Trefferwahrscheinlichkeiten p und Freiwürfen n .

Wer – wie Borovcnik und tausende „braver“ NRW-Abiturienten 2008 – zu b) (1) den von den Aufgabenautoren verlangten Binomialtest der Nullhypothese

$$H_0 : p_A = 0,904 = \frac{288}{288 + 263} \cdot \frac{267}{288} + \frac{263}{288 + 263} \cdot \frac{231}{263}$$

durchführt, nachdem er auf der rechten Seite von Gleichung (1) die entsprechenden empirischen Werte eingesetzt hat, insbesondere also für p_A den empirischen Schätzer $231/263$, hat überhaupt nicht verstanden, dass das Testen einer Nullhypothese – egal ob im Sinne von Fisher oder von Neyman und Pearson – grundsätzlich voraussetzt, dass sich die Nullhypothese unabhängig von den zu ihrer Überprüfung zu erhebenden Daten formulieren lassen muss.

Auch unter der völlig unangemessenen, gleichwohl von Borovcnik akzeptierten Fiktion, nur die Heimtrefferquote sei ein fester Parameter, wäre der in b) (1) geforderte Binomialtest aufgrund der Äquivalenz der Gleichungen (1) und (2) derselbe wie der in b) (2) geforderte Binomialtest, nämlich der der Nullhypothese

$$H_0 : p_A = \frac{267}{288} = 0,927,$$

und er hätte daher selbstverständlich auch dasselbe „signifikante“ Ergebnis erbracht – anders als von

Borovcnik vorgerechnet, anders als von den Aufgabenautoren unterstellt und anders als von hunderten NRW-Lehrern tausenden NRW-Abiturienten gegenüber sanktioniert.

Fazit

Ich hatte dies alles schon in meiner ursprünglichen Kritik an der Nowitzki-Aufgabe ausgeführt. Auch das – sachlich selbstverständlich berechnete – Monitum von Davies gegen die Vermischung von Schätzern und Parametern lief, wenn man es auf die in der Aufgabe skizzierte stochastische Situation zurückbezog, auf Gleiches hinaus: Das Schlimme ist nicht zuvörderst die begriffliche Verwirrung um Schätzer und Parameter im Aufgabentext, sondern dass hier für eine stochastische Situation ein Vorgehen als „vernünftig“ unterstellt wird, das in sich widersprüchlich ist und von daher nicht vernünftig sein kann.

Danach will mir der Verlauf der Diskussion um die Nowitzki-Aufgabe in dieser Zeitschrift als Witz erscheinen, mehr noch als die Aufgabe selbst.

Literatur

- Borovcnik, M. (2009): „Anwendungen“ und Anwendungen – Zentrales Abitur und vergebene Chancen für den Unterricht in Stochastik. *Stochastik in der Schule* 29 (3), 9–18.
- Davies, P. L. (2009): Einige grundsätzliche Überlegungen zu zwei Abituraufgaben. *Stochastik in der Schule* 29 (2), 2–7.
- Diepgen, R. (2008): Kein Witz!?! Zur Nowitzki-Aufgabe im NRW-Zentralabitur 2008. *Stochastik in der Schule* 28 (3), 20–28.

Anschrift des Verfassers

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