

Probleme mit einem Näherungsverfahren im Modell der Normalverteilung

Helmut Wirths, Oldenburg

Zusammenfassung: Die Aufgabe, einen minimalen Stichprobenumfang zu bestimmen, stellt ein häufig gestelltes Problem für Klausuren in der gymnasialen Oberstufe und auch in der schriftlichen Abiturprüfung dar. Sie ist auch deshalb so beliebt, weil man in ihrer Bearbeitung mehr als nur das Abarbeiten von elementaren Routinen sieht, so daß man sie in der Regel als Anforderung höheren Schwierigkeitsgrads ausweist. In der Praxis werden jedoch häufig beim Lösen dieser Aufgabe unabdingbare Überlegungen und unverzichtbare Untersuchungen vernachlässigt.

1. Eine Abituraufgabe aus Bayern

Der fünfte Teil einer Aufgabe aus dem Zentralabitur von Bayern (Abiturprüfung 1996, Wahrscheinlichkeitsrechnung/Statistik III, zitiert nach [6], S. 96-9) lautet:

„Der Sender braucht 400 neue, einwandfreie Magnetbänder. Erfahrungsgemäß sind 2% der gelieferten Bänder schadhafte. Wie viele Bänder müssen bestellt werden, damit mit mindestens 99% Wahrscheinlichkeit wenigstens 400 einwandfreie Bänder darunter sind? Verwenden Sie die Normalverteilung als Näherung.“

Betrachten wir den Lösungsweg, wie er in Ausarbeitungen für eine solche Aufgabenstellung üblicherweise dargestellt wird. Wir erhalten nach einigen Umformungen mit der Φ -Funktion, Berücksichtigung der Stetigkeitskorrektur und nach Übergang zur Umkehrfunktion die Ungleichung (*) $0,98 \cdot n - 2,3264 \cdot \sqrt{0,98 \cdot 0,02} \cdot \sqrt{n} \geq 399,5$. Wir fassen (*) als eine quadratische Ungleichung in der Variablen \sqrt{n} auf, die wir über \sqrt{n} schließlich nach n auflösen. Wir erhalten $n \geq 414,4$. In Musterlösungen finden wir in der Regel als Abschluss der Bearbeitung lediglich eine Ergebnisformulierung wie „Es müssen mindestens 415 Bänder bestellt werden.“

Nun muss eine kritische Überprüfung des Rechenergebnisses erfolgen. Aber die fehlt in der Regel bei den üblichen Musterlösungen. Wir haben ein Näherungsverfahren benutzt und wissen nicht, wie weit das Rechenergebnis von der exakten Lösung entfernt ist. Wir können nur hoffen, nahe an der Lösung zu sein. Außerdem ist solch eine Entscheidung immer mit einem Risiko verbunden, wir können bei einer Bestellung von mindestens 415 Bändern auch weniger als 400 brauchbare erhalten.

2. Eine Abituraufgabe aus Niedersachsen

Bei den mir als Fachberater zur Überprüfung vorgelegten Abituraufgaben aus dem Bereich Stochastik findet sich dieser Aufgabentyp ziemlich häufig. Das im ersten Abschnitt aufgezeigte und zu kritisierende Lösungsverhalten ist auch im dezentralen Abitur in Niedersachsen die Regel. Als besonderen Beleg zitiere ich aus den neuen niedersächsischen „Einheitlichen Prüfungsanforderungen“ (EPA). Teil c der Einzelaufgabe Stochastik „Knallgelb im Hadepark“ (vgl. [5], S. 53/4) lautet:

Damit wir eine sachgemäße Entscheidung treffen können, müssen wir die Sicherheitswahrscheinlichkeit dafür, daß wir unter den genannten Bedingungen mindestens 400 brauchbare Bänder eingekauft haben, kennen. Eine Lösung ohne Angabe dieser konkreten Sicherheitswahrscheinlichkeit ist unvollständig und für den beim Sender für den Einkauf der Bänder Verantwortlichen auch unbrauchbar. Wie man solch eine Sicherheitswahrscheinlichkeit im Modell der Binomialverteilung exakt bestimmen kann, habe ich in [8] beschrieben. Für diese Abituraufgabe erhalten wir mit $n = 415$ und $p = 0,98$, wobei die Zufallsgröße X die Anzahl der brauchbaren Bänder zählt, $P(X \geq 400) = 0,9894$. Wir dürfen nicht darüber erstaunt sein, daß bei einer Bestellung von 415 Bändern die in der Aufgabenstellung geforderte Sicherheit von mindestens 99% nicht erreicht wird, schließlich haben wir die Mindestzahl approximativ bestimmt. Die geforderte Sicherheit ist erst für $n \geq 416$ gegeben. Hätten die für diese Aufgabe in Bayern Verantwortlichen also nicht etwas vorsichtiger formulieren müssen? Darf man überhaupt im Abitur eine Aufgabe stellen und dabei auch noch einen Lösungsweg vorgeben, bei dem man vorher weiß, dass das errechnete Ergebnis nicht Lösung des gestellten Problems ist? Für meine Klausur- und Abituraufgaben lehne ich es ab, so deutlich einen bestimmten Lösungsweg vorzugeben. Auch bei der Sicherheitswahrscheinlichkeit formuliere ich vorsichtiger. Es reicht doch, von „in etwa 99%“ zu reden. Man kann die Angabe einer für den konkreten Problemfall angemessenen Sicherheitswahrscheinlichkeit und die Begründung für die gewählte Größe auch den Lernenden überlassen. Ich mache damit jedenfalls gute Erfahrungen.

„Auf Anfrage bei der Zeitung wurde bestätigt, daß darüber nachgedacht wurde, die Zahl der Abgeordneten durch Nachrücker zu erhöhen, weil in der vergangenen Periode durchschnittlich 20% der Abgeordneten in den Sitzungen nicht anwesend waren. Wie viele Nachrücker müssten gewählt werden, damit mit einer Wahrscheinlichkeit von mindestens 95% mindestens 70 Parlamentarier in den Sitzungen anwesend sind?“

Auch in den niedersächsischen EPA wird unter „Erwartete Lösungswege und Ergebnisse“ eine Modellierung mit Hilfe der Normalverteilung vorgenommen, obwohl die Aufgabenstellung ein solches Vorgehen nicht zwingend vorsieht. Aus dem Ansatz $P(X \geq 70)$, wobei die Zufallsgröße X die Anzahl der anwesenden Parlamentarier zählt, geben die Verfasser nach einigen Umformungen mit der Φ -Funktion, Berücksichtigung der Stetigkeitskorrektur und nach Übergang zur Umkehrfunktion die Ungleichung (+) $\frac{0,8 \cdot n - 69,5}{0,4 \cdot \sqrt{n}} > 1,6449$ an.

Auch hier wird (+) als quadratische Ungleichung in der Variablen \sqrt{n} aufgefasst, die nach \sqrt{n} und anschließend nach n aufgelöst wird. Als Ergebnis wird $n \geq 94,89$ abgedruckt. Die Schlußformulierung der niedersächsischen EPA „95 – 70 = 25 Nachrücker“ erweckt den Eindruck, exakt zu sein. Es kommt darin überhaupt nicht zum Ausdruck, auf welch tönernen Füßen dieses Ergebnis

steht. Es kommt noch schlimmer: „Mindestens 96 Abgeordnete“ lautet die Lösung dieser Aufgabe. Das wird aber nicht erwähnt. Für $n = 96$ und $p = 0,8$ erhalten wir $P(X \geq 70) = 0,965$, während für $n = 95$ diese Sicherheitswahrscheinlichkeit auf drei Nachkommastellen 0,948 beträgt. In der Aufgabenstellung wird aber eine Sicherheit von mindestens 95% gefordert. Bei geschickter Formulierung des Arbeitsauftrags hätte man dieses Dilemma vermeiden können. Außerdem wird noch nachgewiesen, daß die sogenannte Laplace-Bedingung in der Form $\sigma > 3$ für $n > 56$, und damit erst recht für $n > 70$, erfüllt ist. Auf die Laplace-Bedingung wird in Abschnitt 4 eingegangen.

Die oben erwähnte Ungleichung (+) der niedersächsischen EPA ist äquivalent zur Ungleichung (**) $0,8 \cdot n - 1,6449 \cdot \sqrt{0,8 \cdot 0,2} \cdot \sqrt{n} > 69,5$, was man unter Berücksichtigung von $\sqrt{0,8 \cdot 0,2} = 0,4$ und Multiplikation mit dem Nenner leicht überprüfen kann.

3. Ein Lösungsvorschlag

Bei beiden Aufgaben habe ich bewußt die Ungleichungen (*) und (**) besonders erwähnt. Alle Lösungsansätze für diesen Aufgabentyp der Bestimmung einer Maximal- oder Minimalzahl führen auf die in diesen Ungleichungen enthaltene Form, eine geeignete $t \cdot \sigma$ -Umgebung um μ zu betrachten. Dadurch wird folgender Lösungsansatz im Modell der Binomialverteilung nahegelegt: Wir suchen den minimalen Stichprobenumfang n , für den gilt: $\mu - t \cdot \sigma \geq g$ mit $g = 400$ (Bayern) oder $g = 70$ (Niedersachsen). Damit wir eine Lösung für n bestimmen können, benötigen wir einen konkreten Wert für t . Wir müssen t so wählen, daß die Sicherheitswahrscheinlichkeit die Bedingung $P(X \geq 70) \geq 0,95$ (Niedersachsen) beziehungsweise $P(X \geq 400) > 0,99$ (Bayern) erfüllt. Leider ist t bei Binomialverteilungen nicht nur abhängig von der konkreten Sicherheitswahrscheinlichkeit, sondern darüber hinaus auch noch von n und p . Wir wissen, daß wir Binomialverteilungen durch die Standard-Normalverteilung approximieren können. Daher setzen wir für t den Wert ein, den wir bei einer Normalverteilung nehmen würden. Das von mir hier vorgeschlagene Verfahren ist also kein exaktes Lösungsverfahren, sondern auch nur ein Näherungsverfahren. Da wir außer dieser Anleihe aus dem Modell der Normalverteilung ansonsten nur im Modell der Binomialverteilung bleiben, schlage ich die Bezeichnung „Näherungsverfahren im Modell der Binomialverteilung“ vor. Wir wählen $t = 1,6449$ (Niedersachsen) bzw. $t = 2,3264$ (Bayern). Diese Werte für t können wir den üblichen Tabellenwerken (siehe zum Beispiel [2], S. 52 „Quantile der Standardnormalverteilung“, [1], S. 44 „ σ -Bereiche bei normalverteilten Zufallsgrößen“ oder

[4], S. 41 „Schranken bei standardisierter Normalverteilung zur Sicherheitswahrscheinlichkeit $1 - \alpha$ “) entnehmen, mit Hilfe eines zumindest programmierbaren Taschenrechners, eines Taschencomputers, eines Computers oder auf dem üblichen Weg durch Umformungen mit Hilfe der Φ -Funktion berechnen.

Wir müssen bei den hier betrachteten Aufgaben folgende Ungleichungen lösen:

$$- 0,8 \cdot n - 1,6449 \cdot \sqrt{0,8 \cdot 0,2} \cdot \sqrt{n} \geq 70$$

(Niedersachsen)

bzw.

$$- 0,98 \cdot n - 2,3264 \cdot \sqrt{0,98 \cdot 0,02} \cdot \sqrt{n} \geq 400$$

(Bayern).

Wir fassen beide Ungleichungen als quadratische Ungleichungen in der Variablen \sqrt{n} auf, die wir zunächst nach \sqrt{n} und danach nach n auflösen. Wir erhalten als Ergebnis: $n \geq 95,5$ (Niedersachsen) bzw. $n \geq 414,9$ (Bayern). Erst der Einbezug der exakten Sicherheitswahrscheinlichkeiten gibt Aufschluss darüber, daß die Ungleichung $n \geq 96$ die gestellte niedersächsische Aufgabe löst, während erst Anzahlen n mit $n \geq 416$ Lösungen der bayerischen Aufgabe sind. Dieses Näherungsverfahren führt uns etwas besser an die Lösungen heran als das Näherungsverfahren im Modell der Normalverteilung. Es erspart uns aber auch nicht die kritische Überprüfung, ob das durch die Rechnung naheliegende Ergebnis tatsächlich Lösung der gestellten Aufgabe ist. Bei der bayerischen Aufgabe wird deutlich, daß $n = 415$ die gestellte Aufgabe nicht löst.

Am Abschluss der Bearbeitung dieser Aufgaben erfolgt eine Ergebnisdiskussion unter Einbezug der exakten Sicherheitswahrscheinlichkeit. Das gehört unverzichtbar zu einer Bearbeitung mit dazu. Schülerinnen und Schüler müssen bei mir eine begründete Empfehlung aussprechen. Ich mache die Erfahrung, daß absolute Zahlen die Lernenden viel besser zu Antworten anregen und zu Entscheidungen provozieren als relative Zahlen. Daher lasse ich Sicherheitswahrscheinlichkeiten, die ja relative Zahlen sind, auch durch Verhältnisse in natürlichen Zahlen darstellen. Bei der Aufgabe aus Niedersachsen müssen wir erwarten, dass von 1000 Sitzungen, die unter den im Aufgabentext genannten Bedingungen stattfinden, im Mittel rund 52 (bei 95 Parlamentariern) oder in etwa 35 (bei 96 Abgeordneten) schlecht besucht sind, also weniger

4. Abschlussbemerkungen

Wir hätten schon längst über das hier angesprochene Näherungsverfahren im Modell der Normalverteilung nachdenken müssen. Ich denke nicht nur an das Unbehagen, das beim Interpretieren einer Ungleichung wie zum Beispiel $\mu - t \cdot \sigma > 69,5$ in der niedersächsischen Aufgabe entsteht, in der mit Bruchteilen von Abgeordneten gerechnet wird. Ich denke beim Lesen einer solchen Ungleichung vor allem daran, daß wir beim Berechnen der Sicherheitswahrscheinlichkeit $P(E)$ für das Ereignis E einen – wenn auch sehr kleinen – Teil des Gegenereignisses \bar{E} mit einbeziehen. Wir schätzen daher die Sicherheitswahrscheinlichkeit systematisch zu groß ab. Für die hier diskutierte Aufgabenstellung, einen minimalen Stichprobenumfang n zu bestimmen, folgt, daß n systematisch zu klein abgeschätzt wird. Dieser Effekt hätte uns schon längst nachdenklich stimmen müssen. Außerdem hätte uns schon aus reiner Neugier bei der hier diskutierten Problemstellung vor allem die Antwort auf die Frage „Wie groß ist die exakte Sicherheitswahrscheinlichkeit $P(E)$?“ interessieren müssen. Aber gedruckte Musterlösungen wie „Es müssen 95 Abgeordnete sein.“ (vgl. [5]) oder „Es müssen 415 Bänder sein.“ (vgl. [6]) sind schlechte Vorbilder. Diese Antworten sagen nichts darüber aus, auf welch schwankendem Boden sie stehen, wieviel Unsicherheit darin enthalten ist, und erst recht nicht, ob die geforderte Sicherheit tatsächlich erreicht wird. Wir hätten die falschen Vorbilder schon längst korrigieren müssen.

Das in Abschnitt 3 beschriebene Vorgehen ist nicht nur in Leistungskursen, sondern auch in Grundkursen möglich. Es lässt sich unmittelbar nach Erarbeitung einiger grundlegender Eigenschaften von Binomialverteilungen beginnen. In [7] sowie in Abschnitt 7 von [9] habe ich mögliche Unterrichtsinhalte vorgestellt, bei denen das Berechnen von Bereichswahrscheinlichkeiten wesentlicher

als 70 Teilnehmer haben. Diese Möglichkeiten stehen zur Debatte. Oder sollte man vielleicht sogar 97 Abgeordnete empfehlen? In solch einer Diskussion spielen dann auch Argumente wie „Eine ungerade Anzahl von Parlamentariern ist besser, wenn man Patt-Situationen vermeiden will.“ oder „Es lassen sich bei Beachtung der kommunalen Grenzen nur so und so viele in etwa gleich große Wahlkreise einrichten.“ eine Rolle. Damit möchte ich aufzeigen, dass bei Prüfungen zwar der mathematische Beitrag zur Lösung des Problems wesentlich im Vordergrund steht, der in der Realität aber nur einen gewissen Teil zur Entscheidungsfindung beiträgt. Diese Rolle der Mathematik kann und darf bei solch eingekleideten Aufgaben in der Ergebnisdiskussion ruhig deutlich werden.

Bestandteil des Lösungsvorgangs ist. Auch das Berechnen von Höchst- oder Mindestanzahlen ist in diese Unterrichtseinheit eingebettet. Wir können also mit einem wesentlich geringeren theoretischen Aufwand als bisher in der Regel üblich diese auch für Schülerinnen und Schüler interessanten – und auch tatsächlich interessierenden – Problemstellungen in den Unterricht integrieren.

Die sogenannte Laplace-Bedingung (vgl. Abschnitt 2) spielt in diesem Konzept keine Rolle mehr. Das bedaure ich überhaupt nicht. Diese Bedingung muß man im Unterricht vorgeben, sie kann nicht bewiesen werden. Schülerfragen, die auf ein tieferes Verstehen, insbesondere auf den Zusammenhang zwischen Güte der Approximation und der „Schärfe“ der Bedingung ($n \cdot p \cdot q > 9$ wird häufig benutzt, aber auch $n \cdot p \cdot q > 6$ wird zum Beispiel in [3] auf Seite 88 vorgeschlagen) zielen, können nicht zufriedenstellend beantwortet werden.

Bei der Aufgabenformulierung genügt meist eine grobe Angabe bezüglich des Sicherheitsniveaus. Formulierungen wie „große“, „sehr große“ oder „genügend große“ Sicherheitswahrscheinlichkeit reichen aus. Ich habe die Erfahrung gemacht, daß Schülerinnen und Schüler in der Regel sehr verantwortungsbewusst und meist gut begründet ein passendes Sicherheitsniveau wählen. Bei der Ergebnisdiskussion, der in dieser Konzeption ein großes Gewicht zukommt, müssen wir uns zwischen diskreten Sicherheitsniveaus entscheiden. Erst am Ende der Ergebnisdiskussion fällt in der Regel die Entscheidung über das für die Problemstellung angemessene oder anzustrebende Sicherheitsniveau. Es sollte uns dabei auch nicht stören, wenn unter den Lernenden oder zwischen Lernenden und Lehrenden unterschiedliche Auffassungen über das für das jeweilige Problem relevante Sicherheitsbedürfnis bestehen.

Literatur

- [1] Barth, F. u.a. (1981): Stochastik Tabellen. München: Ehrenwirth
- [2] Eggs, H. u.a. (1993): Tafeln zur Stochastik. Frankfurt: Diesterweg
- [3] Krengel, U. (1991): Einführung in die Wahrscheinlichkeitstheorie und Statistik. Vieweg : Braunschweig
- [4] Mühlbauer, P.; Wörle, K. (1983): Tafelwerk zur Stochastik. München: bsv
- [5] Niedersächsisches Kultusministerium (1998): Einheitliche Prüfungsanforderungen in der Abiturprüfung im Lande Niedersachsen - Mathematik. Schroedel: Hannover
- [6] Reimann, S. (1997): Abiturprüfung 1996. Prüfungsaufgaben mit Lösungen. Mathematik Leistungskurs Gymnasium Bayern. Stark: Freising
- [7] Wirths, H. (1996): Schätz- und Prüfverfahren. Mathematik in der Schule Heft 11, S. 596-607, 610-611.
- [8] Wirths, H. (1998): Binomialwahrscheinlichkeiten mit dem Computer. Stochastik in der Schule Heft 1, S. 43-54
- [9] Wirths, H. (1998): Stochastik - Unterrichtsbeispiele und Aufgaben. Oldenburger Vor-Drucke Heft 362/98. Universität Oldenburg: Oldenburg 2. Auflage.

Poisson-Approximation und Kopplung

Peter Eichelsbacher / Helene Worms, Universität Bielefeld

Das Ziel dieser Arbeit ist es, die Poisson-Approximation der Binomial-Verteilung genauer zu untersuchen. Mit einfachen Methoden, die in der Fachwelt *Kopplungsmethoden* genannt werden, kann man genauere Informationen über die Güte der Konvergenz beim klassischen Poissonschen Grenzwertsatz erhalten. Wir diskutieren hier eine sehr einfache Kopplungsmethode. Weiter wollen wir herausarbeiten, daß gerade bei der Poisson-Approximation eine Aussage über die Güte der Annäherung von fundamentaler Bedeutung ist. Das benötigte mathematische Handwerkzeug legt es nahe, in einer Unterrichtseinheit eines Leistungskurses diese Art Gütediskussion führen zu können.

Der vielseitige Physiker und Mathematiker Siméon-Denis Poisson (1781-1840) leitete 1837 in seinem Hauptwerk zur Wahrscheinlichkeitstheorie „Recherches sur la probabilité des jugements, en matière criminelle et en matière civile“ im 81. Kapitel den heute nach ihm benannten Grenzübergang von der Binomial-Verteilung zur Poisson-Verteilung her. (Im Hauptteil dieses Werkes führt er das erstmals so benannte „Gesetz der großen Zahlen“ ein. Es handelt sich dabei um den Grenzwertsatz, der als Spezialfall das von Jakob Bernoulli gefundene Gesetz enthält.) Der Mathematiker von Bortkiewicz machte 1898 auf die gefundene Poisson-Approximation aufmerksam und füllte diese Beobachtung durch eine Vielzahl von Beispielen mit Leben. Sehr bekannt ist wohl das militärische Beispiel der durch Hufschlag getöteten Soldaten, auf das wir hier aber nicht weiter eingehen wollen.

Es sei daran erinnert, daß eine Zufallsgröße X mit der Verteilung

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k q^{n-k}$$

$$(q = 1 - p) \text{ für } k = 0, 1, \dots, n$$

binomialverteilt heißt. Wenn man die Werte der Binomial-Verteilung betrachtet, so stellt man fest, daß bei gleichem Erwartungswert $\alpha = np$ die Werte $b(k; n, p)$ ähnlich sind und daß diese Ähnlichkeit mit wachsendem n zunimmt. Poisson bewies, daß für jeden festen Erwartungswert α eine Grenzverteilung existiert, der die in α übereinstimmenden Binomial-Verteilungen mit wachsendem n zustreben. Da n wächst, entspricht dies der mathematischen Forderung, daß die Folge $np(n)$ gegen eine positive Zahl konvergieren möge, also insbesondere $p(n)$ eine Nullfolge ist. Was bedeutet die Bedingung, daß mit zunehmender Stichprobenzahl die Erfolgswahrscheinlichkeit $p(n)$ klein wird? Beim unabhängigen Münzwurf hat man Interpretationsprobleme! Anders formuliert: bei aller Schönheit des Poissonschen Grenzwertsatzes hat man Probleme, die Bedingung richtig zu verstehen. Man hat streng genommen auch ein Problem bei der mathematischen Modellierung, da man sogenannte Dreiecksschemata einführen müsste. In der folgenden Interpretation der Binomial-Verteilung versteht man die n -Abhängigkeit der Erfolgswahrscheinlichkeit $p(n)$ besser. Man werfe rein zufällig ein Teilchen in ein Volumenelement V und betrachte das Ereignis, daß das Teilchen den Teilbereich $v \subset V$ trifft. Bei n -fach unabhängigem Werfen ist dann die Wahrscheinlichkeit, daß $X = k$ Teilchen v treffen durch $b(k; n, p)$ mit $p = \frac{\text{vol}(v)}{\text{vol}(V)}$ gegeben. Hierbei bezeichne $\text{vol}(A)$ das Volumen