Von der Binomialverteilung zur Normalverteilung
- ein analytischer Ansatz

Dr. Guido Pinkernell

Üblicherweise erfolgt die Einführung in die Normalverteilung als Approximation einer Binomialverteilung $B(n,p)$ für große Stichprobenumfänge n. Die Schülerinnen und Schüler vollziehen diese Approximation in Regel anhand der Histogramme von Binomialverteilungen mit wachsendem n bei gleich bleibendem p, wobei diese bei unveränderter Achseneinteilung breiter und flacher werden. Die vom Lehrer anvisierte Glockenkurve stellt sich nur dann ein, wenn gleichzeitig mit wachsendem n die Diagramme standardisiert werden, so dass diese – horizontal gestaucht, vertikal gestreckt und schließlich übereinander geschoben – das Bild einer Glockenkurve ergeben. Warum diese Transformationen auch sachlich begründet sind, ist den Schülerinnen und Schülern häufig nur schwer zu vermitteln.

Ein bewährter Zugang zur Normalverteilung verzichtet zunächst auf eine stochastische Interpretation der Verteilungskurven, sondern betrachtet sie zunächst als Funktionsgraphen ohne Sachzusammenhang. Die notwendigen Transformationen sind so Gegenstand einer Problemstellung, die eigentlich dem Bereich der Analysis zuzuordnen ist und auch mit üblichen Methoden der Analysis bearbeitet wird. Erst später werden die beteiligten Parameter als Verteilungskennwerte gedeutet.

Im Einzelnen lassen sich die einzelnen Phasen der Unterrichtseinheit wie folgt zusammenfassen: Zunächst werden die Parameterwerte der Binomialverteilung hinsichtlich ihres Einflusses auf die Gestalt des Histogramms untersucht. Anschließend wird die allgemeine Glockenkurve dem Histogramm einer gegebenen Binomialverteilung angepasst. Dann gilt es, diese Parameter hinsichtlich ihrer Bedeutung für die Binomialverteilung zu interpretieren. Und schließlich wird anhand der zur Glockenkurve gehörigen Funktion der Begriff der Dichtefunktion eingeführt.

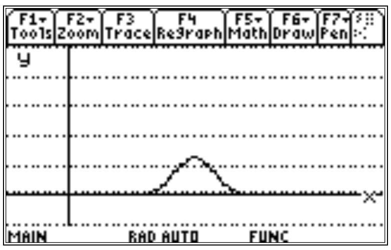

Abb. 1

1. Die Binomialverteilung und ihre Parameter

Die Glockenkurve ist der Graph einer Funktion, die auf ganz IR definiert ist. Dagegen ist der Definitionsbereich der Wahrscheinlichkeitsfunktion einer Binomialverteilung nur eine Teilmenge von IN. Da diese Funktion jetzt Objekt von Parameteruntersuchungen sein soll, muss sie im Rechner für die Eingabe umdefiniert werden:

```
y1=binompdf(40,0.5,round(x,0))
```

wobei die Rechnerfunktion `binompdf(...)` im TI-83/84 Plus/SE vorprogrammiert ist. Beim TI-89 bzw. Voyage™200 muss sie gegebenenfalls nachprogrammiert werden:

```
nCr(n,k)*p^k*(1-p)^(n-k)→binompdf(n,p,k)
```

Die erste Aufgabe hat nun das Ziel, den Einfluss der Verteilungsparameter n und p auf den Graphen einer Binomialverteilung zu erkunden.

Aufgabe 1 *Es sei*

$$P(X=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

die Wahrscheinlichkeitsfunktion einer binomialverteilten Zufallsvariablen X. Untersuchen Sie die Auswirkungen der Parameterwerte n und p auf Position und Gestalt des Graphen und fassen Sie Ihre Beobachtungen mithilfe aussagekräftiger Skizzen zusammen.

Je nach Vorerfahrung der Schüler können folgende Hilfestellungen gegeben werden. Zum einen die Eingabe der Bernoulliformel als Funktionsterm im [Y=]-Editor (Abbildungen 2 und 3). Des weiteren der Hinweis, dass eine Parameteruntersuchung sinnvoller Weise systematisch vorgenommen wird, und zwar, indem man immer nur einen Parameter variiert und die Veränderungen im Graphen beschreibt. Zum Beispiel so:

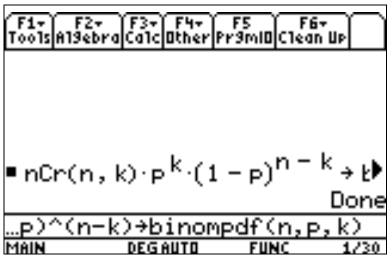

Abb. 2

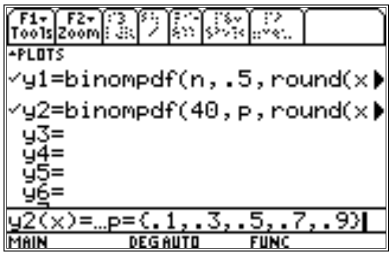

Abb. 3

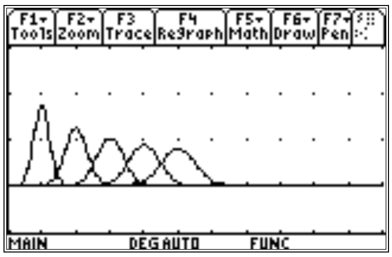

Abb. 4

Mit $p=0{,}5$ und $n\in\{20, 40, 60, 80, 100\}$ erhält man die Abbildung 4. Man erkennt: Alle Kurven sind glockenförmig. Mit wachsendem n verschiebt sich die Kurve nach rechts. Insbesondere liegt die Hochpunktstelle bei $n\cdot 0{,}5$. Außerdem flacht mit wachsendem n die Kurve ab und wird breiter.

Mit $n=40$ und $p\in\{0.1, 0.3, 0.5, 0.7, 0.9\}$ erhält man die Abbildung 5. Man erkennt: Alle Kurven sind glockenförmig. Die Kurve mit $p=0{,}5$ ist die „kleinste" Kurve. Mit wachsendem und mit kleiner werdendem p werden sie höher und schmaler. Insbesondere gehen die Kurven mit $p=0{,}5-k$ und $p=0{,}5+k$ durch Spiegelung an der Geraden mit $x=0{,}5$ paarweise ineinander über.

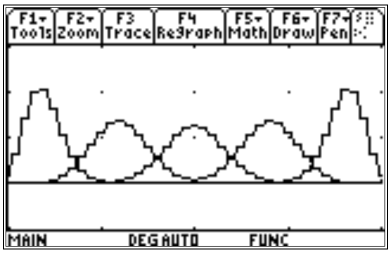

Abb. 5

2. Anpassung der Glockenkurve an die Binomialverteilung

Die zweite Aufgabe hat nun das Ziel, nach einer kurzen Einführung die Glockenkurve so zu verändern, dass sie mit dem Graphen einer gegebenen Binomialverteilung übereinstimmt.

Die Einführung kann auf die glockenförmige Gestalt aller betrachteten Binomialverteilungsgraphen hinweisen. Und wenn man den Ausschnitt sehr groß wählt, dann geht das treppenförmige Glockenprofil in eine knickfreie Glockenkurve über. Eine solche Kurve kann man mithilfe einer Funktion mit „knickfreiem Graphen" approximieren, was die Arbeit mit der Verteilung vereinfachen kann. Die zugehörige Funktionsgleichung lautet:

$$f(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$$

Aufgabe 2 *Passen Sie die Glockenkurve dem Histogramm einer Binomialverteilung an:*

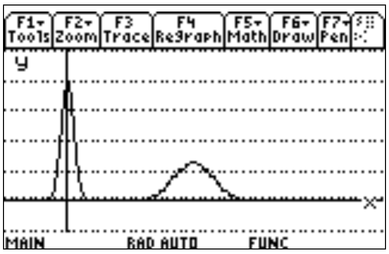

Abb. 6

2.1 *Ändern Sie den Funktionsterm so ab, dass der Graph von f(x) in etwa mit dem Verteilungsgraphen von $B(40{,}0.5)$ übereinstimmt. Und zwar gehen Sie wie folgt vor:*

(a) Verschieben Sie den Graphen von f soweit parallel zur x-Achse, bis die Hochpunktsstelle von f mit der Hochpunktstelle von $B(40{,}0.5)$ übereinstimmt (vgl. Abb. 7).

(b) Stauchen Sie nun den Graphen der neuen Funktion parallel zur y-Achse so, dass beide Hochpunkte übereinstimmen (vgl. Abb. 8).

(c) Strecken Sie nun den Graphen der neuen Funktion parallel zur x-Achse so, dass beide Graphen in Gänze aufeinander zu liegen kommen (vgl. Abb. 9).

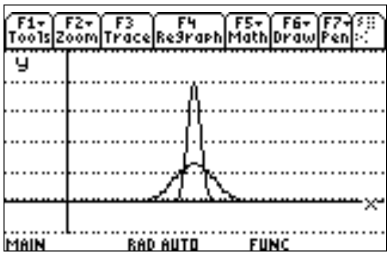

Abb. 7

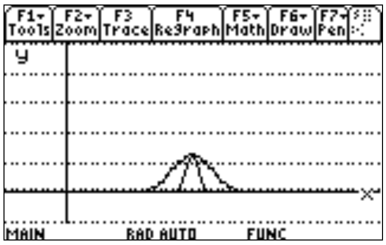

Abb. 8

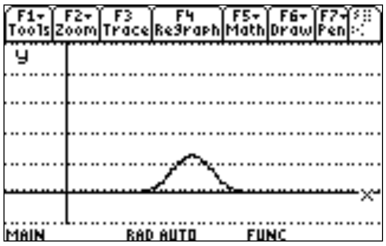

Abb. 9

2.2 *Führen Sie eine entsprechende Transformation für B(80,0.2) durch.*

Auch zu dieser Aufgabe sind einige Hinweise sinnvoll: Man sollte die Schüler darauf aufmerksam machen, dass der TI-83/84 Plus/SE mit der Funktion `normalpdf(...)` über eine Funktion verfügt, deren Graph die benötigte Glockenkurve produziert. Auf die Bedeutung der Bezeichnung „normalpdf" kann dann an geeigneter Stelle hingewiesen werden. Hier ist diese Rechnerfunktion zunächst nur eine nützliche Abkürzung. Auf dem TI-89 sollte sie im Übrigen nachprogrammiert werden (Abbildungen 10 und 11), wobei hier aus technischen Gründen der Name mit `normpdf(...)` kürzer ist[1].

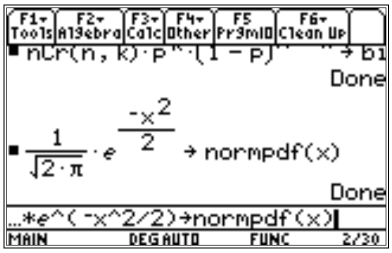

Abb. 10

[1] Natürlich können bei der Nachprogrammierung auf dem TI-89 statt `binompdf(...)` („binomial probability distribution function) bzw. `normalpdf` („normal probability distribution function") ganz andere, kürzere Namen gewählt werden. Hier wurden die gleichen bzw. sehr ähnliche wie die des TI-83 verwendet, damit die Lösungshinweise von Nutzern beider Rechnertypen ohne größere Schwierigkeiten gelesen werden können.

Abb. 11

Die Aufgabenteile in 2.1 können wie folgt bearbeitet werden: Zu (a) ermittelt man durch Ausprobieren eine passende Verschiebung. Wenn Parameteruntersuchungen dieser Art nicht bekannt sind, dann sollte man Hilfen geben, zum Beispiel: „Probiere verschiedene m in `normpdf(x-m)` aus!". Eine geeignete Näherungslösung ist schließlich `normpdf(x-20)`.

Zu (b) kann man einen Stauchungsfaktor von etwa $^1/_3$ als Quotient aus gesuchtem und vorliegendem Maximum in der Regel erwarten. Bei Bedarf ist der folgende Hinweis möglich: „Probiere verschiedene s in `1/s*normpdf(x-20)`" aus. Und zu (c) wird wohl ein Hinweis unumgänglich sein: „Probiere nun verschiedene t in `1/3*normpdf((x-20)/t)`". Es ergibt sich als mögliche Lösung $t \approx 3$.

Zu 2.2: Eine passende Funktion ist hier:
`1/3.6*normpdf((x-40)/3.6)`.

3. Interpretation der Parameter μ und σ

Die gefundenen Parameterwerte $m \approx 20$ und $s = t \approx 3$ haben zunächst nur als Kennwerte der Transformation Bedeutung. In der folgenden Aufgabe werden sie als Kennwerte der Binomialverteilung identifiziert.

Aufgabe 3 *Mit m und $s=t$ haben wir die Glockenkurve an eine gegebene Binomialverteilung angepasst. Diese Werte haben eine besondere Bedeutung für die Binomialverteilung.*

3.1 *Finden Sie das wie folgt heraus:*

(a) Berechnen Sie Erwartungswert, Varianz und Standardabweichung für $B(40{,}0.5)$ und $B(80{,}0.2)$ und beziehen Sie sie auf Ihre Ergebnisse in 2.1 und 2.2.

(b) Informieren Sie sich über den Begriff „Dichtefunktion einer Normalverteilung" und formulieren Sie auf Grundlage von 3.1 den allgemeinen Term einer Dichtefunktion, die $B(n,p)$ approximiert.

3.2 *Fassen Sie zusammen: Welche geometrische Bedeutung haben der Erwartungswert sowie die Standardabweichung für die Dichtefunktion der approximierenden Normalverteilung?*

Zu 3.1: Für $B(40{,}0.5)$ ergeben sich $\mu \approx 20$, $V(x) \approx 8{,}4$ und $\sigma \approx 2{,}9$. Für $B(80{,}0.2)$ ergeben sich analog: $\mu \approx 16$, $V(x) \approx 12{,}8$ und $\sigma \approx 3{,}6$. μ ist unschwer als die Verschiebungslänge zu erkennen. Und σ bestimmt, wie man deutlich sieht, den Stauch- und den Streckfaktor.

Zu 3.2: Die untersuchte Funktion f ist die so genannte Dichtefunktion der approximierenden Normalverteilung. Während eine binomialverteilte Zufallsvariable X_B diskret ist, ist die approximierende normalverteilte Zufallsvariable X_N stetig. Allgemein ergibt sich also für $B(n,p)$ aus den Kennwerten μ und σ als approximierende Dichtefunktion

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Es ist sinnvoll, für die weitere Arbeit mit der Normalverteilung die Funktion `normpdf(...)` im TI-89 um die Parameter μ und σ so zu erweitern, dass sie – wie beim TI-83/84 Plus/SE – als die Dichtefunktion der allgemeinen Normalverteilung definiert ist (Abbildung 12).

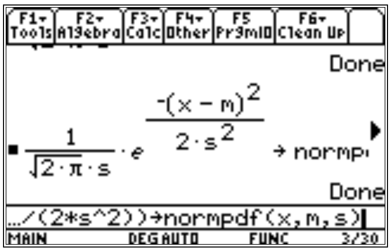

Abb. 12

Und schließlich zu 3.3: μ markiert die Hochpunktstelle des glockenförmigen Graphen und σ misst seine „Breite"[2].

4. Die Dichtefunktion f

Wie Wahrscheinlichkeiten einer binomialverteilten Zufallsvariable berechnet werden ist den Schülern bekannt. Wie dasselbe mit der approximierenden Dichtefunktion erreicht werden kann, ist in der folgenden Aufgabe herauszufinden:

Aufgabe 4 *Die Dichtefunktion nähert nicht nur das graphische Profil einer Binomialverteilung. Sie erlaubt auch eine näherungsweise Berechnung der Wahrscheinlichkeiten.*

4.1 *Berechnen Sie zunächst einige Wahrscheinlichkeiten der Binomialverteilung:*

(a) Zeichnen Sie das Histogramm einer $B(10{,}0.3)$-verteilten Zufallsvariablen X_B in das abgebildete Koordinatensystem.

(b) Berechnen Sie $P(X_B=1)$, $P(X_B=2)$, $P(X_B=3)$ und $P(1 \le X_B \le 3)$ und schraffieren Sie im Histogramm den Bereich, an dem man $P(1 \le X_B \le 3)$ ablesen kann.

4.2 *Ermitteln Sie diese Wahrscheinlichkeiten nun mithilfe der zugehörigen Dichtefunktion f:*

(a) Berechnen Sie den Term der Dichtefunktion f mithilfe von Erwartungswert und Standardabweichung und zeichnen Sie den Graphen der Dichtefunktion f über das Histogramm.

(b) Entwickeln Sie eine Idee, wie man mithilfe von f die Wahrscheinlichkeit $P(1 \le X_N \le 3)$ berechnen kann. Schlagen Sie gegebenenfalls in Ihrer Formelsammlung nach, wie die Wahrscheinlichkeit für eine stetige Zufallsvariable definiert ist und erläutern Sie die Gültigkeit der Formel unter Bezugnahme auf Ihre bisherigen Ergebnisse.

---

[2] Genauer ist 2σ der horizontale Abstand zwischen den beiden Wendepunkten links und rechts vom Hochpunkt. Dies kann in einer weiteren Aufgabe untersucht werden.

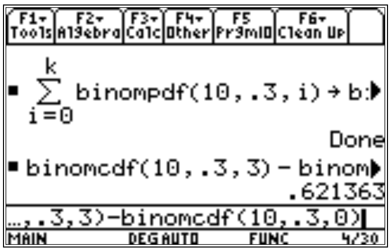

Abb. 13

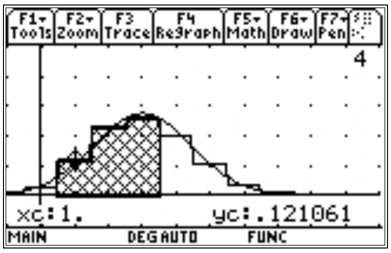

Abb. 14

Zu 4.1: Es sind $P(X_B=1)\approx 0.121$, $P(X_B=2)\approx 0.233$, und $P(X_B=3)\approx 0.267$. Damit ist $P(1\leq X_B\leq 3)\approx 0.621$. Hier bietet sich übrigens eine Gelegenheit, das Konzept der kumulativen Verteilungsfunktion einzuführen. Der TI-83 enthält diese unter dem Namen `binomcdf(...)`, auf dem TI-89 lässt sie sich nachprogrammieren (vgl. Abb. 13). Abbildung 14 zeigt die verlangte Zeichnung mit Schraffur. Schüler werden nicht zwingenderweise eine Fläche markieren. Der Lehrer sollte aber ggf. darauf hinweisen, dass in einem Histogramm Häufigkeits- bzw. Wahrscheinlichkeitswerte als Flächeninhalte dargestellt werden und eine „flächige Markierung" einfordern, damit anschließend der Bezug zum Integral der Dichtefunktion hergestellt werden kann. Hier würde man also rechnen:

$$P(1\leq X_B\leq 3)\approx 0{,}121\cdot 1+0{,}233\cdot 1+0{,}267\cdot 1=0{,}621$$

Zu 4.2 Die zu $B(10{,}0.3)$ gehörigen Kennwerte lauten $\mu\approx 3$ und $\sigma\approx 1{,}45$. Damit ist

$$f(x)=\frac{1}{1{,}45\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-3)^2}{2\cdot 1{,}45^2}}.$$

Die Schraffierung aus Aufgabe 4.1 erinnert an die Ober- und Untersummen als Näherungen des bestimmten Integrals

$$\int_1^3 f(x)\,dx.$$

Und in der Tat kann man einer Formelsammlung entnehmen, dass für die Zufallsvariable X_N und der zugehörigen Dichtefunktion f gilt:

$$P(1\leq X_N\leq 3)=\int_1^3 f(x)\,dx.$$

Literatur

Brandt, D.: *Mathematik unterrichten mit dem TI-83 und TI-83 Plus in Klassenstufe 11-13. Teil III - Stochastik*. Texas Instruments 2004

Juen, H.: *Binomialverteilung, Normalverteilung. Aufgaben für Schüler. Workshopmaterialien*. Seminar für Projektleiter des Österreichischen CAS-Projekts. Ossiach 1999

Autor:

Dr. Guido Pinkernell, Lingen (D)
Gymnasium Johanneum Lingen
guido.pinkernell@gmx.de