

# Der verwirrende Siegeszug des Histogramms in deutsche Klassenzimmer: Sind Stabdiagramme tot?

Wir plädieren dafür, diskrete Verteilungen wie früher ausschließlich mithilfe von Stabdiagrammen zu veranschaulichen, damit solche Verteilungen nicht mit Flächen und folglich mit stetigen Zufallsvariablen assoziiert werden. Das mittlerweile als Darstellungsmittel dominierende Histogramm sollte der Veranschaulichung einer nach Gruppierung von Daten entstandenen empirischen Häufigkeitsverteilung sowie der Approximation einer diskreten durch eine stetige Verteilung wie beim zentralen Grenzwertsatz vorbehalten bleiben.

## 1 Einleitung

Im gymnasialen Stochastikunterricht werden sowohl diskrete als auch stetige Zufallsvariablen behandelt. Der gewaltige konzeptionelle Unterschied zwischen beiden besteht darin, dass diskrete Zufallsvariablen einzelne Werte mit Wahrscheinlichkeitsmassen belegen, wohingegen stetige Zufallsvariablen eine Dichte besitzen und die Wahrscheinlichkeit, in ein Intervall zu fallen, durch die Fläche zwischen der  $x$ -Achse und dem Schaubild der Dichte über dem Intervall gegeben ist.

Insbesondere nimmt eine stetige Zufallsvariable jeden vorgegebenen Wert mit der Wahrscheinlichkeit null an. Man sollte also mit einer diskreten Zufallsvariablen einzelne *Punkte* und mit einer stetigen Zufallsvariablen *Flächen* assoziieren. Dieser Assoziation wurde in früheren Schulbüchern wie z.B. [Barth und Haller (1992)], [Glaser et al. (1990)] und [Feuerpfeil et al. (1994)] dadurch Rechnung getragen, dass man diskrete Verteilungen mithilfe von *Stabdiagrammen* veranschaulichte. Überraschenderweise ist der Begriff des Stabdiagramms jedoch in neuerer Zeit zum Teil vollständig verschwunden, siehe z.B. [Strick (2008)], [Lergemüller et al. (2012)] und [Freudigmann et al. (2016)].

Natürlich fragt man sich, was an die Stelle des Stabdiagramms getreten ist, denn dem Konzept der Dichte für eine stetige Zufallsvariable muss ja etwas entgegengesetzt werden, was die „Punktmassen“ betont. Die Antwort auf diese Frage ist das *Histogramm*, welches in den letzten 10 Jahren einen unaufhaltsamen Siegeszug angetreten hat. Da das Histogramm Wahrscheinlichkeiten mithilfe von *Flächen* darstellt, ist es für die Veranschaulichung diskreter Verteilungen denkbar ungeeignet, und sein Wortschöpfer Karl Pearson hatte sich auch etwas ganz anderes dabei gedacht. Mit diesem Aufsatz möchten wir eine Lanze für das Stabdiagramm brechen, ohne das Histogramm aufzugeben. Beide Begriffe sollten gut verstanden und auseinandergehalten werden.

## 2 Das Histogramm

Es war Karl Pearson, der Begründer der Mathematischen Statistik, der in der letzten seiner *Gresham Lectures* am 20.11.1891 den Begriff *Histogramm* einfürte (siehe [Magnello (1996), S. 48]), und er schrieb dazu [Pearson (1895), S. 399]:

*Introduced by the writer in his lectures on statistics as a common term of graphical representation, i.e. by columns marking as areas the frequency corresponding to the range of their base.*

Das Histogramm ist somit ein Mittel zur grafischen Darstellung empirischer Häufigkeitsverteilungen. Soll für einen Datensatz von  $n$  nicht notwendig paarweise verschiedenen reellen Zahlen  $x_1, \dots, x_n$  ein Histogramm erstellt werden, so bildet man zunächst  $s$  disjunkte Klassen  $[a_1, a_2), [a_2, a_3), \dots, [a_s, a_{s+1})$  in Form „direkt aneinander stoßender“ Intervalle, deren Vereinigung alle  $x_1, \dots, x_n$  enthält. Dann errichtet man über jedem Intervall  $[a_j, a_{j+1})$  als Grundseite ein Rechteck, dessen Fläche gleich dem relativen Anteil der in das Intervall  $[a_j, a_{j+1})$  fallenden Daten ist. Als Beispiel betrachten wir 102 Messwerte eines vor kurzem in mehreren 11. Klassen durchgeführten Handspannenexperimentes. Dabei wurden die Längen der linken und rechten Handspanne gemessen sowie zusätzliche Merkmale wie das Geschlecht notiert. Diese Daten dienen dem zweiten Autor in der Unterrichtseinheit „Beschreibende Statistik“ als roter Faden. Die Lernenden waren motiviert, stellten echte Forschungsfragen und entdeckten anhand der nachstehenden, selbst erhobenen Daten den Nutzen verschiedener Darstellungsformen und Kenngrößen.

|      |      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 23.5 | 22.0 | 19.9 | 19.5 | 20.8 | 20.6 | 16.9 | 21.6 | 21.8 | 20.7 | 22.0 | 23.8 |
| 22.4 | 23.2 | 24.1 | 22.6 | 18.2 | 21.6 | 20.5 | 20.3 | 20.4 | 24.5 | 24.0 | 22.6 |
| 21.2 | 21.1 | 23.1 | 20.0 | 19.5 | 22.6 | 18.8 | 23.5 | 20.2 | 21.0 | 19.2 | 20.9 |
| 23.0 | 21.8 | 19.3 | 19.9 | 20.5 | 20.9 | 20.2 | 23.0 | 17.3 | 20.6 | 20.3 | 23.5 |
| 19.3 | 21.0 | 19.5 | 23.0 | 22.4 | 20.0 | 18.7 | 21.4 | 21.6 | 18.6 | 20.6 | 20.8 |
| 18.2 | 21.0 | 18.2 | 19.0 | 21.0 | 18.0 | 21.5 | 22.1 | 20.5 | 22.5 | 25.5 | 20.2 |
| 20.5 | 18.1 | 22.4 | 19.8 | 17.0 | 20.6 | 20.0 | 18.7 | 23.4 | 22.2 | 21.2 | 19.4 |
| 17.5 | 20.4 | 21.3 | 18.5 | 18.6 | 22.6 | 23.0 | 22.0 | 22.7 | 21.2 | 19.2 | 19.0 |
| 21.4 | 20.0 | 18.5 | 20.8 | 17.7 | 21.5 |      |      |      |      |      |      |

Wählen wir  $a_1 = 16, a_2 = 17.3, a_3 = 18.6, a_4 = 19.9, a_5 = 21.2, a_6 = 22.5, a_7 = 23.8, a_8 = 25.1, a_9 = 26.4$ , also  $s = 8$  Klassen, so ergibt sich das in Abb. 1 dargestellte Histogramm.

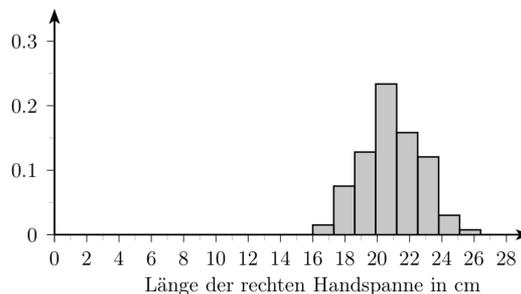


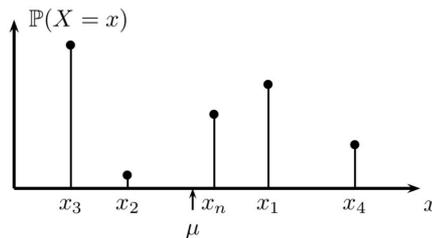
Abb.1: Histogramm (Länge der rechten Handspanne in cm)

Obwohl bei diesem Histogramm gleiche Klassenbreiten gewählt wurden, können die Grundseiten der einzelnen Rechtecke eines Histogramms im Allgemeinen unterschiedlich lang sein. So kompensiert man etwa eine geringere Daten-Dichte in den Extrembereichen der Stichprobe, also der kleinsten und größten Werte, dadurch, dass man die Grundseiten der Rechtecke in diesen Bereichen verlängert, um nicht zu wenige Daten in einer Klasse zu haben. Natürlich stellt sich bei der Anfertigung eines Histogramms grundsätzlich auch die Frage nach der Anzahl  $s$  der Klassen. Es ist plausibel, dass  $s$  mit wachsendem Stichprobenumfang  $n$  ebenfalls zunehmen sollte. Eine grobe Faustregel besagt,  $s \approx \sqrt{n}$  zu setzen. Die Anzahl  $s$  der Klassen sollte also von der Größenordnung  $\sqrt{n}$  sein. Damit zusammenhängende Fragen betreffen die nichtparametrische Dichteschätzung, siehe Abschnitt 5.

Abschließend sei gesagt, dass der Ursprung des Wortes *Histogramm* wohl vom griechischen „*histos*“ (Mast) kommt, siehe [Rufilanchas (2017), S. 32], und nichts mit dem Wort *history* zu tun hat.

### 3 Das Stabdiagramm

Eine Zufallsvariable  $X$  heißt *diskret* (verteilt), wenn sie mit Wahrscheinlichkeit 1 endlich viele oder abzählbar-unendlich viele verschiedene Werte annimmt. Der erste Fall tritt etwa für eine binomialverteilte Zufallsvariable ein, der zweite für eine Poisson-Verteilung. Die mit einer diskret verteilten Zufallsvariablen einhergehende anschauliche Vorstellung ist die einer „Massen-Verteilung“. Nimmt  $X$  die Werte  $x_1, \dots, x_n$  mit positiven Wahrscheinlichkeiten an, so erhält der „Punkt“  $x_j$  auf der reellen Achse die dort angebrachte Wahrscheinlichkeit (smasse)  $\mathbb{P}(X = x_j), j = 1, \dots, n$ .



**Abb. 2:** Stabdiagramm einer diskreten Verteilung mit Erwartungswert als Schwerpunkt

Abb. 2 zeigt das Stabdiagramm der Verteilung einer diskreten Zufallsvariablen. Mit eingezeichnet ist der sich aus der Gleichgewichtsbedingung  $\sum_{j=1}^n (x_j - \mu) \cdot \mathbb{P}(X = x_j) = 0$  ergebende Erwartungswert  $\mu$  als physikalischer Schwerpunkt dieser Massen-Verteilung, wenn die  $x$ -Achse als gewichtslos angenommen wird.

Die formal auf der ganzen reellen Achse definierte Funktion, die jedem  $x$  die Wahrscheinlichkeit  $\mathbb{P}(X = x)$  zuordnet, nennt man oft *Wahrscheinlichkeitsfunktion* (engl.: *probability mass function*) von  $X$ . Für die grafische Darstellung einer solchen diskreten Verteilung war früher der Begriff *Stabdiagramm* üblich. So findet man etwa das linke Bild (mit sehr schmalen Stäben) in Abb. 3 in [Barth und Haller (1992), S. 173], und das rechte (mit etwas breiteren Stäben) in [Glaser et al. (1990), S. 10]. Beide Bilder machen hinreichend deutlich, dass die Zufallsvariable  $X$  ausschließlich die Werte 2,3, ...,12 annimmt. Diesem

Umstand wird auch dadurch Rechnung getragen, dass man die reelle Achse mit dem für natürliche Zahlen vorbehaltenen Symbol  $k$  und nicht mit  $x$  beschriftet. Auch [Feuerpfeil et al. (1994), S. 74, S. 111] verwenden den Begriff Stabdiagramm für eine Darstellung wie in Abb. 3.

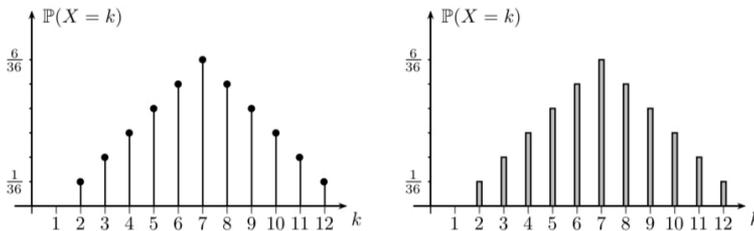


Abb. 3: Stabdiagramme der Verteilung der Augensumme beim zweifachen Würfelwurf

Interessanterweise ging man im Laufe der Zeit bei Darstellungen diskreter Verteilungen in Schulbüchern dazu über, aus den Stabdiagrammen *Histogramme* zu machen und den Begriff Stabdiagramm wie etwa in [Strick (2008)], [Lergenmüller et al. (2012)] und [Freudigmann et al. (2016)] sogar vollständig wegzulassen. So findet sich auf S. 42 in [Strick (2008)] die in Abb. 4 gegebene Darstellung für die Verteilung der Augensumme beim zweifachen Würfelwurf, und man liest dazu:

*Wahrscheinlichkeitsverteilungen können durch **Histogramme** veranschaulicht werden: Auf der horizontalen Achse werden die Werte der Zufallsgröße  $X$  abgetragen, auf der Vertikalen die zugehörigen Wahrscheinlichkeiten  $\mathbb{P}(X = k)$ . Die Rechtecke haben alle die Breite 1, die Höhe der Rechtecke entspricht den Wahrscheinlichkeiten. Die Maßzahl der Fläche der gesamten Figur ist gleich 1.*

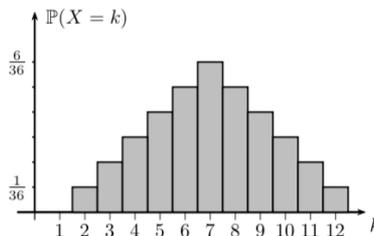


Abb. 4: Histogramm der Verteilung der Augensumme beim zweifachen Würfelwurf

Die apodiktische Sprechweise, dass die „Rechtecke alle die Breite 1 haben“ - wobei jedoch vorher allgemein von „Wahrscheinlichkeitsverteilungen“ gesprochen wird - suggeriert, dass es nur diskrete Wahrscheinlichkeitsverteilungen gäbe, die auf äquidistanten, sich um jeweils 1 unterscheidenden Werten konzentriert sind, was natürlich zu einer eklatanten Fehlvorstellung führt. An dieser Stelle sind [Barth und Haller (1992)] konzeptionell noch deutlich umfassender. Dort werden auf S. 166/167 auch Rechtecke mit allgemeiner Breite  $\Delta x$  thematisiert, um bei beliebigen diskreten Verteilungen Überlappungen von Rechtecken zu vermeiden. Bei gleicher Rechteckbreite sind dann die Höhen der Rechtecke proportional zu den jeweiligen Wahrscheinlichkeiten. Da die Rechtecke nicht notwendig aneinander

stoßen müssen, sind bei diesem allgemeineren Ansatz auch sehr schmale Rechtecke möglich, die dann wie im rechten Bild von Abb. 3 als Stabdiagramme bezeichnet werden können.

#### 4 Von der Binomialverteilung zur Normalverteilung – jetzt werden Flächen nötig

Die Notwendigkeit, Punktmassen „zu Flächen verschmieren zu müssen“, zeigt sich bei der Motivation des zentralen Grenzwertsatzes von de Moivre-Laplace für die Binomialverteilung. Dieser Grenzwertsatz betrifft die Wahrscheinlichkeit, dass eine *standardisierte* binomialverteilte Zufallsvariable Werte in einem vorgegebenen Intervall annimmt. Besitzt die Zufallsvariable  $S_n$  eine Binomialverteilung mit den Parametern  $n$  und  $p$ , wobei  $0 < p < 1$ , so gilt  $\mathbb{P}(S_n = k) = b_n(k, p)$  für  $k = 0, 1, \dots, n$ . Dabei wurde kurz

$$b_n(k, p) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

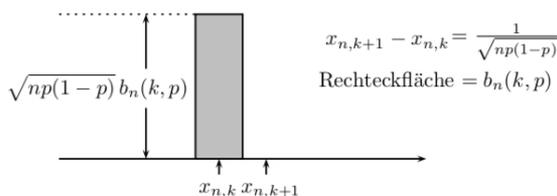
gesetzt. Da Erwartungswert ( $np$ ) und Varianz ( $np(1-p)$ ) von  $S_n$  bei wachsendem  $n$  über alle Grenzen streben, wandert der Schwerpunkt dieser diskreten Verteilung für größer werdendes  $n$  anschaulich gesprochen „immer weiter nach rechts“. Zudem wird das zugehörige Stabdiagramm „immer flacher“. Beide Effekte treten nicht auf, wenn man die Binomialverteilung *standardisiert*, also die Zufallsvariable

$$\tilde{S}_n := \frac{S_n - np}{\sqrt{np(1-p)}}$$

betrachtet. Diese nimmt die Werte

$$x_{n,k} := \frac{k - np}{\sqrt{np(1-p)}}, \quad k = 0, 1, \dots, n,$$

an, und sie besitzt den Erwartungswert null und die Varianz eins. Nach Konstruktion gilt  $S_n = k$  genau dann, wenn die Realisierung von  $\tilde{S}_n$  den Wert  $x_{n,k}$  ergibt.



**Abb. 5:** Transformation der in  $x_{n,k}$  angebrachten Punktmasse zu einer Rechteckfläche

„Verschmiert“ man jetzt die im Punkt  $x_{n,k}$  angebrachte Wahrscheinlichkeitsmasse zur Fläche eines Rechtecks, dessen Grundseite den Mittelpunkt  $x_{n,k}$  und die Länge  $x_{n,k+1} - x_{n,k} = 1/\sqrt{np(1-p)}$  sowie die Höhe  $\sqrt{np(1-p)} \cdot b_n(k, p)$  besitzt, so hat dieses Rechteck gerade die Fläche  $b_n(k, p)$ , siehe Abb. 5.

Um eine Einsicht zu vermitteln, wie die Aussage des zentralen Grenzwertsatzes von de Moivre-Laplace zu verstehen ist, wurde dieser Übergang von Punktmassen zu Flächen früher in Schulbüchern durchgeführt. So schreiben etwa [Feuerpfeil et al. (1994)] auf S. 129: „Für die folgenden Ausführungen eignet sich die Histogrammdarstellung der Binomialverteilung besser als das Stabdiagramm“. Noch detaillierter gehen [Barth und Haller (1992), S. 276ff.], auf die Approximation der Binomialverteilung durch die Normalverteilung ein. Im Zusammenhang mit der Standardisierung heißt es etwa auf S. 279: „Im Gegensatz zum Histogramm, wo die Rechteckfläche ein Maß für die Wahrscheinlichkeit ist, verändert sich das Stabdiagramm bei der Standardisierung nicht in der Höhe, da hier die Stablänge ein Maß für die Wahrscheinlichkeit ist“. Wir sehen also, dass früher begrifflich streng zwischen Stabdiagramm und Histogramm unterschieden wurde.

Konstruiert man für jedes  $k = 0, 1, \dots, n$  ein Rechteck wie in Abb. 5, so entsteht ein Histogramm der standardisierten Binomialverteilung. Abb. 6 zeigt anhand des Falls  $p = 0.3$ , welcher Effekt sich einstellt, wenn man so vorgeht und den Parameter  $n$  vergrößert: Wo hingegen das Histogramm der standardisierten Binomialverteilung in den Fällen  $n = 5$  und  $n = 20$  (Bild links bzw. rechts oben) in Bezug auf die vertikale Achse noch eine prägnante Asymmetrie aufweist, erscheint es schon für den Fall  $n = 100$  deutlich symmetrischer. Im Fall  $n = 100$  ist zusätzlich der Graph der durch

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty,$$

definierten und oft als *Gaußsche Glockenkurve* bezeichneten Dichte der Standardnormalverteilung eingezeichnet.

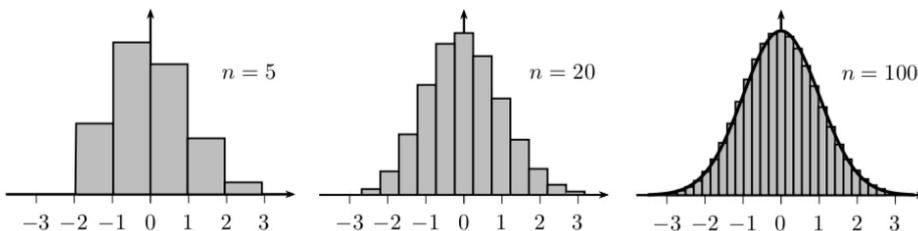


Abb. 6: Histogramme standardisierter Binomialverteilungen für  $p = 0,3$

Die Aussage des zentralen Grenzwertsatzes von de Moivre-Laplace besagt nun, dass für jede Wahl von  $a$  und  $b$  mit  $a < b$  die Wahrscheinlichkeit, dass die standardisierte Zufallsvariable  $\tilde{S}_n$  einen Wert im Intervall  $[a, b]$  annimmt, also das Ereignis  $a \leq \tilde{S}_n \leq b$  eintritt, beim Grenzübergang  $n \rightarrow \infty$  gegen das Integral  $\int_a^b \varphi(x) dx$  konvergiert.

Wegen  $\int_{-1}^1 \varphi(x) dx \approx 0.682$  und

$$\begin{aligned} -1 \leq \tilde{S}_n \leq 1 &\Leftrightarrow -1 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq 1 \\ &\Leftrightarrow np - \sqrt{np(1-p)} \leq S_n \leq np + \sqrt{np(1-p)} \end{aligned}$$

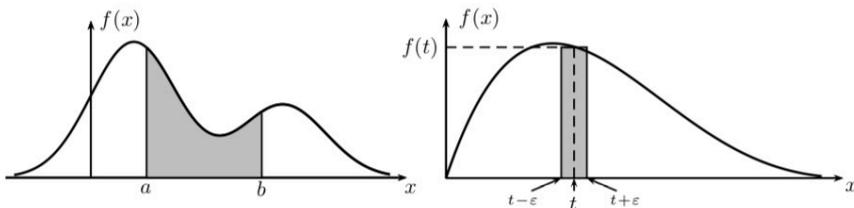
folgt dann insbesondere: Die Wahrscheinlichkeit, dass bei großem  $n$  die Trefferanzahl in einer Bernoulli-Kette der Länge  $n$  vom Erwartungswert  $np$  um höchstens  $\sqrt{np(1-p)}$  abweicht, ist ca. 68 %. Eine solche Aussage findet sich heute als reines Rezept ohne jegliche Hintergrundinformation, siehe z.B. [Freudigmann et al. (2016), S. 147]. Eine derartige Vorgehensweise wirft natürlich grundsätzliche Fragen nach dem Sinn von Mathematikunterricht auf, siehe [Henze (2018b)].

### 5 Stetige Zufallsvariablen, Dichteschätzung

Stetige Zufallsvariablen sind Gegenstand der gymnasialen Kursstufe, siehe z.B. [Dürr et al. (2017), S. 289ff.]. Dabei werden die Gleichverteilung auf einem Intervall sowie die Normalverteilung und die Exponentialverteilung genauer betrachtet. Das grundlegende, mit einer stetigen (genauer: *absolut stetigen*) Zufallsvariablen verknüpfte Konzept ist das einer Wahrscheinlichkeitsdichte oder kurz *Dichte*. Für schulische Zwecke ist eine Dichte eine nichtnegative, bis auf endlich viele Stellen stetige, nichtnegative Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  mit  $\int_{-\infty}^{\infty} f(x)dx = 1$ . Eine Zufallsvariable  $X$  heißt *stetig verteilt* mit Dichte  $f$ , wenn für jedes Intervall  $[a, b]$  mit  $a < b$  gilt:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx .$$

Wohingegen die „Intervall-Wahrscheinlichkeit“  $\mathbb{P}(a \leq X \leq b)$  im Fall einer diskreten Zufallsvariablen gleich der Summe der „Punkt-Wahrscheinlichkeiten“  $\mathbb{P}(X = x_j)$  über alle  $x_j$  mit  $a \leq x_j \leq b$  ist, kann  $\mathbb{P}(a \leq X \leq b)$  jetzt als Fläche zwischen der  $x$ -Achse und dem Graphen von  $f$  über dem Intervall  $[a, b]$  gedeutet werden, siehe Abb. 7 links.



**Abb. 7:** Wahrscheinlichkeit  $\mathbb{P}(a \leq X \leq b)$  als Flächeninhalt (links) und Interpretation des Dichtebegriffs (rechts)

Da eine beliebige reelle Zahl  $x$  für jedes  $\epsilon > 0$  im Intervall  $[x - \epsilon, x + \epsilon]$  enthalten ist, also die Teilmengenbeziehung  $\{x\} \subset [x - \epsilon, x + \epsilon]$  besteht, gilt  $\mathbb{P}(X = x) = 0$ . Im Gegensatz zu einer diskreten Zufallsvariablen nimmt also eine stetige Zufallsvariable jeden einzelnen Wert  $x$  mit Wahrscheinlichkeit null an.

Ist  $t$  eine beliebige Stelle, an der die Dichte  $f$  stetig ist, so kann man den Wert  $f(t)$  wie folgt interpretieren. Für kleines  $\epsilon > 0$  gilt

$$\mathbb{P}(t - \epsilon \leq X \leq t + \epsilon) = \int_{t-\epsilon}^{t+\epsilon} f(x)dx \approx 2\epsilon f(t)$$

und somit (vgl. Abb. 7 rechts)

$$f(t) \approx \frac{1}{2\epsilon} \mathbb{P}(t - \epsilon \leq X \leq t + \epsilon).$$

Lässt man  $\epsilon$  gegen null streben, so gilt

$$f(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbb{P}(t - \epsilon \leq X \leq t + \epsilon)}{2\epsilon}. \quad (1)$$

Der Wert der Dichte an der Stelle  $t$  lässt sich also als Grenzwert von Wahrscheinlichkeitsmasse in einem kleinen Intervall mit Mittelpunkt  $t$ , dividiert durch die Länge dieses Intervalls, deuten, wenn sich das Intervall auf den Punkt  $t$  „zusammenzieht“. Diese Beziehung legt folgenden Ansatz zur Schätzung von  $f(t)$  nahe, wenn die Dichte  $f$  unbekannt ist und man Realisierungen stochastisch unabhängiger Zufallsvariablen  $X_1, \dots, X_n$  mit gleicher Dichte  $f$  beobachten kann: Schreiben wir allgemein  $\mathbf{1}\{A\}$  für die Indikatorfunktion eines Ereignisses  $A$ , also diejenige Zufallsvariable, die die Werte 1 bzw. 0 annimmt, je nachdem, ob das Ereignis  $A$  eintritt oder nicht (siehe z.B. [Henze (2018a), S. 12]), so kann man die im Zähler von (1) auftretende Wahrscheinlichkeit durch den zufälligen relativen Anteil

$$\frac{1}{n} \sum_{j=1}^n \mathbf{1}\{t - \epsilon \leq X_j \leq t + \epsilon\}$$

derjenigen unter den  $X_1, \dots, X_n$ , die in das Intervall  $[t - \epsilon, t + \epsilon]$  fallen, schätzen. Da mit Wahrscheinlichkeit 1 keines der  $X_j$  den Wert  $t$  annimmt, konvergiert dieser zufällig relative Anteil für  $\epsilon \rightarrow 0$  gegen null. Wir können also nur hoffen, den Wert  $f(t)$  bei wachsendem  $n$  immer genauer schätzen zu können, wenn wir einen Schätzer der Form

$$\hat{f}_n(t) = \frac{1}{2\epsilon_n} \cdot \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{t - \epsilon_n \leq X_j \leq t + \epsilon_n\} \quad (2)$$

ins Auge fassen. Dabei ist  $(\epsilon_n)$  eine von  $n$  abhängende Folge positiver reeller Zahlen, die geeignet gegen null konvergiert (wir werden sehen, was das heißt!). Von der Verteilung der Zufallsvariablen  $\hat{f}_n(t)$  erwarten wir, dass sie sich bei wachsendem  $n$  immer stärker um den Wert  $f(t)$  konzentriert. Diese Forderung können wir durch die Limesbeziehungen

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}_n(t)) = f(t), \quad (3)$$

$$\lim_{n \rightarrow \infty} \mathbb{V}(\hat{f}_n(t)) = 0 \quad (4)$$

präzisieren. Da die in der Definition von  $\hat{f}_n(t)$  stehende Indikatorsumme die Binomialverteilung  $\text{Bin}(n, p_n(t))$  mit

$$p_n(t) = \int_{t-\epsilon_n}^{t+\epsilon_n} f(x) dx$$

besitzt, folgt mit der allgemeinen Rechenregel  $\mathbb{E}(aY) = a\mathbb{E}(Y)$  für Erwartungswerte

$$\mathbb{E}(\hat{f}_n(t)) = \frac{1}{2\epsilon_n} \cdot \frac{1}{n} \cdot np_n(t) = \frac{p_n(t)}{2\epsilon_n}$$

und damit (3). Der Schätzer  $\hat{f}_n(t)$  ist also *asymptotisch erwartungstreu* für  $f(t)$ . Mit der Rechenregel  $\mathbb{V}(aY) = a^2\mathbb{V}(Y)$  für die Varianz ergibt sich weiter

$$\mathbb{V}(\hat{f}_n(t)) = \frac{1}{4\epsilon_n^2} \cdot \frac{1}{n^2} \cdot np_n(t)(1 - p_n(t)) = \frac{1}{2\epsilon_n n} \cdot \frac{p_n(t)}{2\epsilon_n} \cdot (1 - p_n(t)).$$

Da der zweite Faktor auf der rechten Seite gegen  $f(t)$  und der dritte gegen 1 konvergieren, muss  $\epsilon_n n \rightarrow \infty$  gelten, damit (4) erfüllt ist. Die Folge  $(\epsilon_n)$  darf also nicht zu schnell gegen null konvergieren. Mit der Wahl  $\epsilon_n = n^{-1/2}$  gelten sowohl (3) als auch (4).

Dass das durch (2) gegebene Schätzverfahren etwas mit dem „Verschmieren von Punktmassen zu Flächen“ zu tun hat, wird ersichtlich, wenn man die auch als *Rechteckkern* bezeichnete Funktion  $K(y) := 1/2$ , falls  $|y| \leq 1$ , und  $K(y) := 0$ , sonst, einführt. Damit wird (2) zu

$$\hat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n \frac{1}{\epsilon_n} K\left(\frac{X_j - t}{\epsilon_n}\right).$$

Da die Funktion  $t \mapsto K((X_j - t)/\epsilon_n)$  die Dichte der (stetigen) Gleichverteilung auf dem Intervall  $[X_j - \epsilon_n, X_j + \epsilon_n]$  ist, kann man diese Dichte-Schätzmethode bei Vorliegen von Realisierungen  $x_1, \dots, x_n$  von  $X_1, \dots, X_n$  so beschreiben: Man transformiert die durch die jeweiligen Punktmassen  $1/n$  an den Stellen  $x_1, \dots, x_n$  vorliegende empirische Häufigkeitsverteilung in eine Dichte, indem man die Dichten der Gleichverteilungen auf den Intervallen  $[x_j - \epsilon_n, x_j + \epsilon_n]$ ,  $j = 1, \dots, n$ , überlagert und mit dem Faktor  $1/n$  multipliziert, damit die Funktion  $\hat{f}_n(t)$ ,  $t \in \mathbb{R}$ , eine Dichte wird.

Der Rechteckkern kann erstaunlich einfach mithilfe von GeoGebra realisiert werden. So lautet der Befehl für die Definition der Funktion  $\hat{f}_n$

$$\text{„}f(x) = 1/(2\epsilon \cdot n) \cdot \text{Summe}(\text{Wenn}(x - \epsilon \leq \text{Element}(\text{Daten}, i) \leq x + \epsilon, 1, 0), i, 1, n)\text{“},$$

was einfach eine 1:1-Übersetzung von (2) ist. Die Liste *Daten* enthält die 102 Messwerte. Durch Variation von  $\epsilon$  verändert sich die Schätzung. Dieser Umstand kann gewinnbringend im Unterricht demonstriert werden. Abb. 8 zeigt das Histogramm der Handspannendaten zusammen mit dem Graphen des Dichteschätzers  $\hat{f}_n$ , wobei  $\epsilon = 1,2$  gewählt wurde.

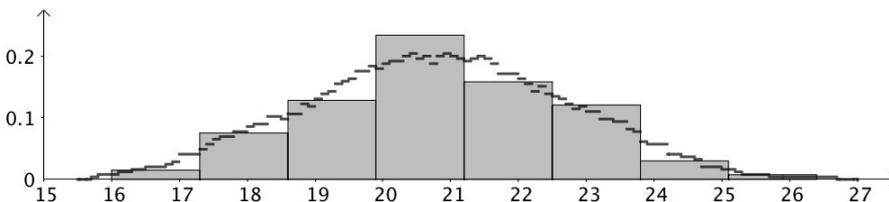


Abb. 8: Dichteschätzung anhand der Handspannendaten,  $\epsilon = 1,2; n = 102$

## 6 Fazit

Mathematikunterricht lebt von Konzepten; vermittelt werden aber leider zu oft nur Rezepte. Grundlegende Konzepte der Stochastik sind diskrete und stetige Zufallsvariablen. Eine diskrete Zufallsvariable nimmt nur endlich oder abzählbar-unendlich viele Werte an und liefert so eine „Punktmassen-Verteilung“. Für solche Verteilungen sind Stabdiagramme das natürliche Darstellungsmittel. Spätestens, wenn man die Verteilung einer Zufallsvariablen veranschaulichen will, die die Werte -0,2, 0,5 und 2 annimmt (eine solche Situation könnte aufgrund eines Glücksspiels entstehen, bei dem man 20 Cent verlieren, aber auch 50 Cent oder 2 Euro gewinnen kann), erkennt man, dass ein Histogramm mit Rechteckbreite 1 völlig unsinnig ist. Diskrete Verteilungen haben nichts mit Flächen zu tun! Hier wird Verwirrung gestiftet, und es werden falsche Fährten gelegt. Dass Stabdiagramme aus Schulbüchern verschwunden sind, liegt wohl in erster Linie daran, dass im Unterricht momentan fast ausschließlich die Binomialverteilung behandelt wird.

Im Gegensatz zu einer diskreten Zufallsvariablen nimmt eine stetige Zufallsvariable jeden einzelnen Wert mit der Wahrscheinlichkeit null an. Das damit einhergehende Konzept ist das einer Dichte, womit *Wahrscheinlichkeiten von Intervallen* zu Flächen werden. Das Histogramm sollte der Darstellung einer empirischen Häufigkeitsverteilung vorbehalten bleiben. Die Notwendigkeit, Punktmassen „zu Flächen verschmieren zu müssen“, zeigt sich insbesondere bei der Verdeutlichung des zentralen Grenzwertsatzes von de Moivre-Laplace für die Binomialverteilung und beim Problem der Schätzung einer Dichte.

### Literatur

- Barth, F., und Haller, R. (1992): Stochastik Leistungskurs. Ehrenwirth Verlag, München.
- Baum, M. et al. (2007): Lambacher Schweizer 5 Mathematik für Gymnasien, 1. Auflage, Verlag E. Klett, Stuttgart.
- Dürr, R. et al. (2017): Lambacher Schweizer. Mathematik für Gymnasien. Kursstufe. Verlag E. Klett, Stuttgart.
- Feuerpfeil, J. et al (1994): Praktische Stochastik, Bayerischer Schulbuch-Verlag München, 2. Auflage.
- Freudigmann, H. et al. (2016): Lambacher Schweizer. Mathematik für Gymnasien 10, Verlag E. Klett, Stuttgart.
- Glaser, H., et al. (1990): Sigma: Grundkurs Stochastik, 1. Auflage: Verlag E. Klett, Stuttgart.
- Henze, N. (2018a): Stochastik für Einsteiger. 12. Auflage: Verlag Springer Spektrum. Heidelberg.
- Henze, N. (2018b): Verständnisorientierter gymnasialer Stochastikunterricht - quo vadis? In: *Stochastik in der Schule* 38 (2018) 3, 12-23.
- Lergenmüller, A. et al. (2012): Mathematik Neue Wege. Arbeitsbuch für Gymnasien. Stochastik. Bildungshaus Schulbuchverlage, Braunschweig.
- Magnello, M. Eileen (1996): Karl Pearson's Gresham Lectures: W.F.R. Weldon, Speciation and the origin of Pearsonian statistics. In: *The British Journal for the History of Science*, Vol. 29(1), 43-63.
- Pearson, K. (1895): X. Contributions to the Mathematical Theory of Evolution - II. Skew variation in Homogeneous Material. In: *Philos. Trans. of the Royal Statist. Soc. A*, 186 (1895), 343-424.
- Rufilanchas, D.R. (2017): On the origin of Karl Pearson's term „histogram“. In: *Estadística Española*, 59, número 192, 29-35.
- Strick, H.-K. (2008): Einführung in die beurteilende Statistik. Bildungshaus Schulbuchverlage, Braunschweig.