stik in der Schule, Heft 3, Band 2 (1982)

Eine weitere Möglichkeit, wie der Begriff des Erwartungswertes auch eingeführt werden kann*

von Alan W.Sykes

(übersetzt von W.D.Heller)

Der Erwartungswert wird in fast allen Statistikbüchern in der gleichen Vorgehensweise eingeführt. Hat man zunächst den Stichprobenmittelwert über die beobachteten Häufigkeiten definiert, wird man in natürlicher Weise den Erwartungswert einer diskreten Zufallsvariablen X, die die Werte $x_1, x_2, \ldots$ mit den Wahrscheinlichkeiten $p_1, p_2, \ldots$ annimmt, als

$$E(X) = p_1x_1 + p_2x_2 + \ldots \qquad (1)$$

definieren.

Stetige Zufallsvariablen X, deren Verteilung durch eine Dichtefunktion $f_X(x)$ gegeben sind, benötigen eine "eigene" Definition von E(X), die lautet:

$$E(X) = \int x f_X(x)\,dx. \qquad (2)$$

Natürlich liegt es nahe, Formel (2) als stetige Version von (1) zu erklären. Im folgenden soll eine Einführung des Erwartungswertbegriffs betrachtet werden, die in der gängigen obigen Art sowohl den diskreten Fall (1) als auch den kontinuierlichen Fall (2) zusammenfaßt (aber nicht ersetzt!). Diese Herleitung ist *einfach*, weil sehr gut zu veranschaulichen, darüber hinaus aber auf diskrete *und* stetige Zufallsvariablenanwendbar.

Die neue Methode

Die Zufallsvariable X habe die Verteilungsfunktion $F_X(x) = P(X \le x)$. In Abbildung 1 wird die Verteilungsfunktion einer stetigen Zufallsvariablen X graphisch dargestellt.

* Originaltitel in 'TEACHING STATISTICS' (1981) Heft 3, Band 3 'An Alternative Approach to the Mean'

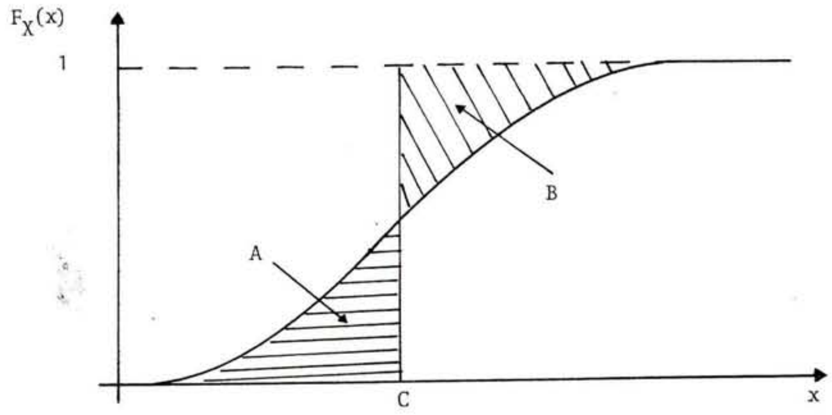


Abb. 1: Verteilungsfunktion $F_X(x)$ einer stetigen Zufallsvariablen X

Ein Punkt (c,o) werde auf der x-Achse beliebig gewählt; es sei dann A die Fläche zwischen x=c, $F_X(x)$ und der x-Achse, die Fläche B wird von x=c, $F_X(x)$ und y=1 begrenzt. Unter der Voraussetzung, daß die beiden Flächen A und B endlich sind, existiert E(X) und es gilt:

$$E(X) = c + B-A. \qquad (3)$$

Festzustellen ist:

(i) die Definition von E(X) ist von c unabhängig.

(ii) E(X) ist genau dann gleich c, wenn A = B ist.

(iii) Ist X eine positivwertige Zufallsvariable, dann ist E(X) gleich der Fläche zwischen $F_X(x)$ und y=1.

Die Möglichkeit, den Erwartungswert nach (3) zu definieren, wird in den meisten Statistikbüchern nicht genutzt, während im Buch von Feller ([1]) zumindest die Eigenschaft (iii) von (3) angegeben wird. Im folgenden sollen einige Aspekte der Definition des Erwartungswerts durch (3) kurz angesprochen werden.

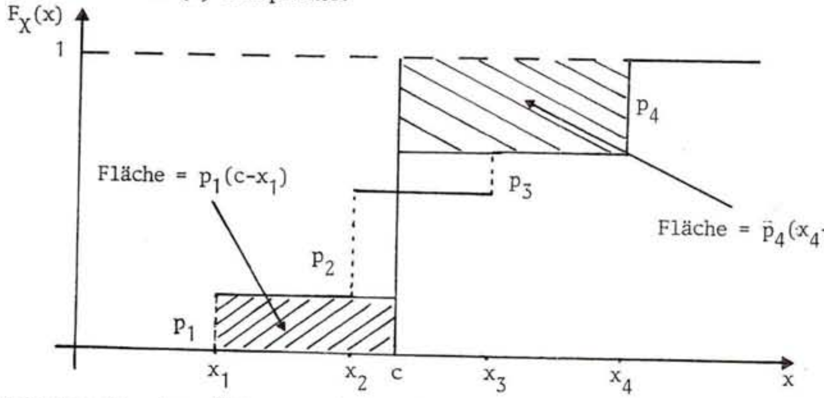
Äquivalenz der Formeln (1) und (3)

Eine Zufallsvariable X nehme die Werte x_1, x_2, x_3, x_4 mit den Wahrscheinlichkeiten p_1, p_2, p_3 und p_4 $(p_1+p_2+p_3+p_4=1)$ an. Die Verteilungsfunktion $F_X(x)$ ist eine Treppenfunktion, die an den Stellen x_i die Sprunghöhen p_i besitzt. Wendet man (3) wie oben gezeigt an, so folgt:

$$A = p_1(c-x_1) + p_2(c-x_2)$$

$$B = p_3(x_3-c) + p_4(x_4-c)$$

Nach einigen kleinen Umformungen ergibt sich $c+B-A=p_1x_1+p_2x_2+p_3x_3+p_4x_4$, was der Definition (1) entspricht.



Beispiel für eine diskrete Zufallsvariable, deren Werte gleiche Abstände haben

X sei eine diskrete Zufallsvariable, deren Verteilung durch die beiden ersten Zeilen von Tabelle 1 gegeben sei. Wird nun für c ein Wert des Wertebereichs von X gewählt - z.B. c=6 - so muß als nächstes die Wahrscheinlichkeitsmasse, die oberhalb und unterhalb von c liegt, aufsummiert werden (Zeile 3 in Tabelle 1). Wiederholt man diesen Vorgang, so ergibt sich Zeile 4 der Tabelle 1. In beiden Fällen werden die Wahrscheinlichkeiten, die zu dem gewählten c-Wert gehören, nicht berücksichtigt. Die in Tabelle 1 eingekreisten Werte werden nun mit 2 multipliziert um A und B zu erhalten. Somit ist $E(X) = 6+2\cdot(0.7-0.4) = 6.6$.

Tabelle 1

Wertebereich von X	2	4	6	8	10
$P(X=x_i)$	0.1	0.2	0.3	0.1	0.3
Summierte Wahrscheinlichkeiten (Stufe 1)	0.1	0.3	-	0.4	0.3
Summierte Wahrscheinlichkeiten (Stufe 2)	0.1	$\boxed{0.4}$		$\boxed{0.7}$	0.3

Wird für c anstelle von 6 der Wert 8 gewählt, ändern sich die Zeilen 3 und 4 von Tabelle 1 folgendermaßen:

Wertebereich von X	2	4	6	8	10
Summierte Wahrscheinlichkeiten (Stufe 1)	0.1	0.3	0.6	·	0.3
Summierte Wahrscheinlichkeiten (Stufe 2)	0.1	0.4	1.0	·	0.3

und $E(X)$ ergibt sich zu $E(X) = 8 +2(0.3-1.0) = 6.6$.

Gruppierte Werte, wobei die Klassenbreite konstant ist

Durch Tabelle 2 ist eine Häufigkeitsverteilung gegeben, deren Stichprobenwerte in 5 Klassen gleicher Breite zusammengefaßt werden. Die die Klassen repräsentierenden Werte (die Klassenmitten) sowie die relativen Häufigkeiten stimmen mit den Werten von Tabelle 1 überein. Der Stichprobenmittelwert der Daten wird analog dem Vorgehen in Tabelle 1 bestimmt, die Vorgehensweise sei aber abweichend von Tabelle 1 nun in Tabelle 2 für absolute Häufigkeiten dargestellt.

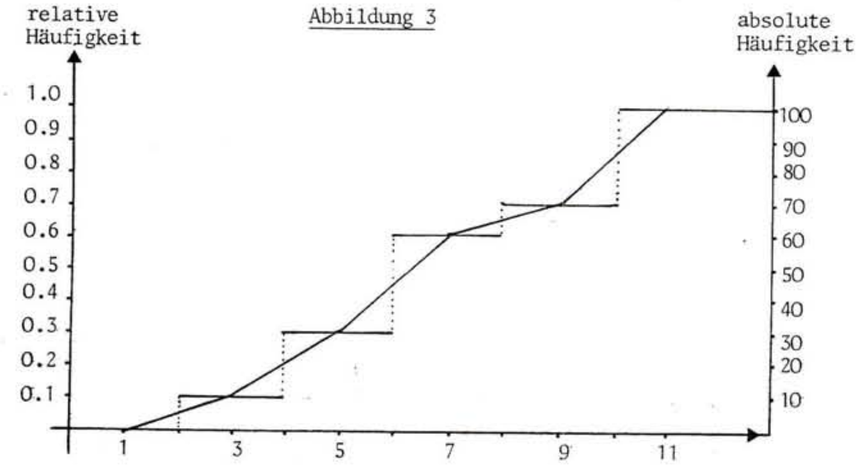
Tabelle 2

Klassen [ )	1 bis 3	3 bis 5	5 bis 7	7 bis 9	9 bis 11
Klassenmitte	2	4	6	8	10
absolute Häufigkeit	10	20	30	10	30
summierte absolute Häufigk. (Stufe 1)	10	30		40	30
summierte absolute Häufigk. (Stufe 2)	10	$\boxed{40}$		$\boxed{70}$	30

$$\text{Stichprobenmittelwert} = 6 + \frac{2}{100}(70-40) = 6.6.$$

Da die Urdaten nicht vorliegen, kann die Berechnung des Stichprobenmittelwertes nur unter gewissen Annahmen durchgeführt werden. Die Berechnungen in Tabelle 2 setzen voraus, daß die Wahrscheinlichkeitssumme von jeder Klasse auf die Klassenmitte konzentriert ist - die Stichprobenverteilungsfunktion (oder auch empirische Verteilungsfunktion) also eine Treppenfunktion ist (Abb. 3). Realistischer wäre es anzunehmen, daß die Werte innerhalb einer Klasse gleichverteilt sind, so daß sich als Stichprobenverteilungsfunktion der in Abb. 3 dargestellte Polygonzug ergibt. Die Anwendung unserer neuen Definition (3) auf beide Verteilungsfunktionen ergibt das gleiche Resultat (die relevanten Flächen A und B sind anhand von Abbildung 3 zu berechnen, wobei die Berechnung der Flächen A und B leichter sein wird als die der entsprechenden Flächen des Polygonzuges).

Abbildung 3



Berechnung von E(X) für eine "gemischte Zufallsvariable"

Die Formeln (1) und (2) können nur auf Zufallsvariablen angewendet werden, die entweder diskret oder stetig sind. Nicht angewendet werden können sie für Zufallsvariablen, die weder diskret noch stetig sind und die also weder eine Treppenfunktion noch eine stetige Funktion als Verteilungsfunktion besitzen. Die in Formel (3) benötigten Flächen A und B hingegen können immer berechnet werden. Sei etwa folgende Verteilungsfunktion $F_X(x)$ der Zufallsvariablen X gegeben:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ x & \text{für } 0 \leq x < \frac{1}{4} \\ \frac{1}{2}x + \frac{1}{2} & \text{für } \frac{1}{4} \leq x \leq 1 \\ 1 & \text{für } x > 1 \end{cases}$$

Der Graph von $F_X(x)$ weist eine Unstetigkeitsstelle bei $x = \frac{1}{4}$ auf. Nimmt man deswegen $c=\frac{1}{4}$ und wendet (3) an, so läßt sich über die Dreiecksflächen A und B mit den Flächeninhalten von 1/32 und 9/64 E(X) zu

$$E(X) = \frac{1}{4} + 9/64 - 1/32 = 23/64$$

berechnen.

Die Markoffsche Ungleichung

Die Tschebyscheffsche Ungleichung, ein Korollar der Markoffschen Ungleichung, kann für den Beweis des schwachen Gesetzes der großen Zahlen verwendet werden. Obwohl dieses Gebiet der Wahrscheinlichkeitsrechnung sicherlich nicht allzu tief im Stochastik-Schulunterricht behandelt werden dürfte bzw. könnte, soll kurz auf eine Anwendung von Formel (3) zur Begründung der Markoffschen Ungleichung eingegangen werden.
Für eine positivwertige Zufallsvariable X, die den Erwartungswert E(X) besitzt, lautet die Markoffsche Ungleichung:

$$P(X \geq \lambda) \leq \frac{E(X)}{\lambda}, \quad \forall \lambda > 0.$$ [1]

Diese Ungleichung ergibt sich in natürlicher Weise aus Formel (3): E(X) ist die durch $F_X(x)$, y = 1 und x = 0 begrenzte Fläche, $\lambda \cdot P(X>\lambda)$ ist hingegen die Fläche eines Rechtecks, das in der E(X) charakterisierenden Fläche enthalten ist.

Zusammenfassung

Es ist nicht schwierig die Grenzen der in dieser Arbeit betrachteten Definitionsmöglichkeit von E(X) aufzuweisen - E(X) einer Poissonverteilung über (3) zu bestimmen wäre ein "tödliches" Unterfangen. Des weiteren benötigt man zur Berech-

[1] Die Markoffsche Ungleichung gilt noch für einen allgemeineren Fall:
$P(|X| \geq \lambda) \leq \frac{E(|X|^r)}{\lambda^r}$, wobei X eine beliebige Zufallsvariable und λ, r>o sind.
Als Spezialfall erhält man dann für r=2 und Y=X-E(X) die Tschebyscheffsche Ungleichung.

nung von $E(X^2)$ nach (3) die Verteilungsfunktion von X^2. Nichts desto weniger kann die Formel (3) und mit ihr ihre Anwendungen als elementar, anschaulich und als ein erfolgreicher Schritt bei der Zusammenfassung der Formeln (1) und (2) für den diskreten bzw. stetigen Fall eingestuft werden. Sind dies nicht genug gute Gründe, um (3) in elementaren Stochastikbüchern zu erwähnen?

Literatur

[1] Feller, W. (1950): Introduction to Probability and its Applications (vol.2), S. 148, Wiley.