Prof. Dr. Klaus Eckhardt

Wahrscheinlichkeitsdichte

veröffentlicht im Internet unter aufgabomat.de

Gliederung:

1 Diskrete und stetige Zufallsvariablen

Eine Zufallsvariable X[1] heißt **diskret**, wenn sie nur endlich viele oder abzählbar unendlich viele Werte annimmt. „Abzählbar unendlich" heißt, dass die Werte mit den natürlichen Zahlen durchnummeriert werden können. In ihrer symbolischen Schreibweise werden solche Werte daher üblicherweise mit einem Index versehen und beispielsweise als x_k oder x_i bezeichnet. Eine diskrete Zufallsvariable ist vollständig beschrieben, wenn ihre **Wahrscheinlichkeitsfunktion** bekannt ist. Die Werte $f(x_k)$ der Wahrscheinlichkeitsfunktion entsprechen den Wahrscheinlichkeiten, dass die diskrete Zufallsvariable ihre Werte x_k annimmt:

$$f(x_k) = P(X = x_k). \tag{1}$$

P ist eine Abkürzung des lateinischen Wortes „probabilitas" für die Wahrscheinlichkeit.

Eine **stetige** Zufallsvariable kann dagegen innerhalb eines gewissen Bereiches jeden reellen Wert annehmen. Dies sind, wie man in der Mathematik sagt, überabzählbar unendlich viele Werte, d. h. ihre Anzahl übersteigt die Anzahl der natürlichen Zahlen - auch wenn letztere unendlich groß ist. Es gibt verschiedene Ausmaße von Unendlichkeit.

Um eine stetige Zufallsvariable vollständig zu beschreiben, muss ihre Wahrscheinlichkeitsdichtefunktion identifiziert werden. Die Ausführungen im vorliegenden Skript dienen dazu, die Wahrscheinlichkeitsdichtefunktion anhand einer Analogie einzuführen, derjenige zur bekannteren Massendichte.

2 Wahrscheinlichkeitsdichte

Stellen Sie sich eine Ansammlung von Materie in einem Bereich V des dreidimensionalen Raums vor. Materie besitzt Masse, und so lässt sich die Materieansammlung dadurch beschreiben, wie sich die Masse in diesem Raum verteilt. Einem Punkt lässt sich keine Masse zuordnen, denn ein Punkt hat keine räumliche Ausdehnung, aber die Masse von Volumenelementen kann erfasst und angegeben werden.

[1] Anmerkung zu einer in der Statistik üblichen Konvention, der auch im vorliegenden Dokument gefolgt wird: Eine Zufallsvariable wird in der Statistik symbolisch mit einem Großbuchstaben bezeichnet, z. B. mit X. Als Symbol für einen Wert der Variable wird der zugehörige Kleinbuchstabe verwendet, z. B. x.

Nehmen wir an, der Raumbereich V wird in eine endliche Zahl n von quaderförmigen Volumenelementen ΔV_k mit den Kantenlängen Δx_k, Δy_k und Δz_k unterteilt (Abbildung 1):

$$\Delta V_k = \Delta x_k \, \Delta y_k \, \Delta z_k.$$

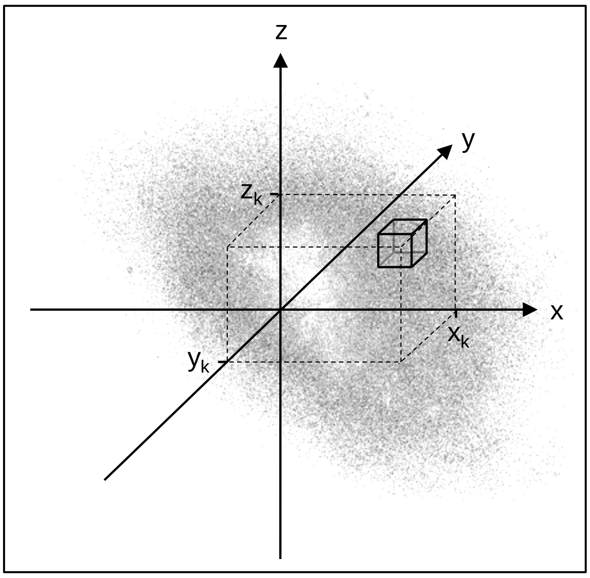

Abbildung 1: Volumenelement mit den Koordinaten x_k, y_k und z_k in einem materieerfüllten Raumbereich.

Die Position der Volumenelemente werde durch die Koordinaten x_k, y_k und z_k angegeben. Hat das Volumenelement ΔV_k die Masse Δm_k, so lässt sich für dieses Volumenelement die mittlere Massendichte $\rho(x_k,y_k,z_k) = \Delta m_k/\Delta V_k$ bestimmen. Ist umgekehrt die Massendichte $\rho(x_k,y_k,z_k)$ gegeben, so lässt sich die Masse des Volumenelements berechnen als

$$\begin{aligned} \Delta m_k &= \rho(x_k,y_k,z_k)\, \Delta V_k \\ &= \rho(x_k,y_k,z_k)\, \Delta x_k \, \Delta y_k \, \Delta z_k\,. \end{aligned}$$

Die Gesamtmasse im Raumbereich V ist dann

$$m(V) = \sum_{k=1}^{n} \rho(x_k, y_k, z_k)\, \Delta x_k \, \Delta y_k \, \Delta z_k\ .$$

Allerdings ist dies eine relativ grobe Beschreibung der Materieansammlung, denn die Massendichte $\rho(x_k,y_k,z_k)$ ist ein über das Raumelement $\Delta V_k = \Delta x_k \, \Delta y_k \, \Delta z_k$ gemittelter Wert. Die räumlich differenzierten, lokalen Verhältnisse werden umso genauer beschrieben, je kleiner ΔV_k ist. Werden die einzelnen Volumenelemente kleiner, so steigt zugleich ihre Anzahl, d. h. man muss die Anzahl n der Volumenelemente erhöhen und letztlich zum Grenzfall $n \to \infty$ übergehen:

$$m(V) = \lim_{n\to\infty} \sum_{k=1}^{n} \rho(x_k, y_k, z_k)\, \Delta x_k \, \Delta y_k \, \Delta z_k\ .$$

Für den Grenzwert auf der rechten Seite dieser Gleichung ist in der Mathematik eine abkürzende Schreibweise eingeführt worden:

$$m(V) = \iiint_V \rho(x, y, z)\, dx\, dy\, dz\,. \qquad (2)$$

In Worten: Die Masse m im Raumbereich V wird berechnet, indem die lokal variierenden Massendichte ρ über diesen Raumbereich integriert wird.

Jetzt zur Statistik. Weder verteilt sich Materie gleichmäßig über den gesamten Raum, noch nimmt eine Zufallsvariable sämtliche Werte in der Menge der reellen Zahlen mit ein und derselben Wahrscheinlichkeit an.

Beispiel: Die Länge eines Stabes wird gemessen. Zur Messung steht nur ein Geodreieck mit Millimeterskala zur Verfügung, das mehrfach angelegt werden muss. Daraus resultiert eine Ungenauigkeit, die dazu führt, dass die Messwerte bei wiederholter Messung in der Größenordnung von Millimetern variieren (Tabelle 1).

Messung Nr.	x (cm)	Messung Nr.	x (cm)
1	100,2	7	100,0
2	100,1	8	99,7
3	99,9	9	100,1
4	100,1	10	100,4
5	99,9	11	100,0
6	100,3	12	100,3

Tabelle 1: Beispiel für zwölf zufällig streuende Werte der gemessenen Länge X eines Stabes.

Der Stab ist rund einen Meter lang. Entsprechend häufen sich die Messwerte bei etwa einem Meter bzw. gibt es eine erhöhte Wahrscheinlichkeit dafür, dass die Werte der Zufallsvariable X bei etwa einem Meter liegen. Es ist weder zu erwarten, dass plötzlich ein Messwert von zwei Metern auftritt, noch, dass X negative Werte annimmt.

So, wie eine Materieansammlung mit einer bestimmten Verteilung von Masse im dreidimensionalen Raum verbunden ist, ist eine Zufallsvariable mit einer bestimmten Verteilung von Wahrscheinlichkeit in der Menge der reellen Zahlen verbunden. Die Betrachtung wird dadurch erleichtert, dass sich die Menge der reellen Zahlen als eindimensionaler Raum auffassen lässt: Um die Position einer Zahl in |R zu beschreiben, reicht eine Koordinatenachse aus, der Zahlenstrahl. Identifizieren Sie also

- den dreidimensionalen Raum mit der Menge |R der reellen Zahlen
- Punkte des Raums mit einzelnen Zahlen $x \in$ |R
- den Raumbereich V mit einem Zahlenbereich, einem Intervall [a; b].

Eine Materieansammlung lässt sich durch die im Raum variable Massendichte ρ(x,y,z) beschreiben. Die Masse in einem Raumbereich V wird berechnet, indem man die Massendichte über V integriert (Gleichung 2). Analog lässt sich eine Zufallsvariable durch ihre in der Menge der reellen Zahlen variable Wahrscheinlichkeitsdichte f(x) beschreiben. Die Wahrscheinlichkeit, dass die Zufallsvariable X einen Wert in einem Intervall [a; b] annimmt, wird berechnet, indem man die Wahrscheinlichkeitsdichte über das Intervall [a; b] integriert:

$$P(a \le X \le b) = \int_a^b f(x)\,dx \,. \qquad (3)$$

Die Funktion f(x) wird als **Wahrscheinlichkeitsdichtefunktion** bezeichnet.

Da ein bestimmtes Integral, wie es auf der rechten Seite der Gleichung 3 steht, einer Fläche unter einer Funktionskurve entspricht, lässt sich auch sagen: Wahrscheinlichkeiten, mit denen Zufallsvariablen

Werte in vorgegebenen Intervallen annehmen, entsprechen Flächen unter der zugehörigen Wahrscheinlichkeitsdichtefunktion.

Kennt man die Wahrscheinlichkeitsdichtefunktion, durch die eine Zufallsvariable X beschrieben wird, lässt sich gemäß Gleichung 3 prinzipiell jede Wahrscheinlichkeit für die betreffende Zufallsvariable berechnen. Ein zentrales Ziel der Statistik ist es daher, zu identifizieren, durch welche Wahrscheinlichkeitsdichtefunktion sich eine gegebene Zufallsvariable beschreiben lässt.

Hier abschließend die zuvor analog angeführten Größen noch einmal in der Übersicht:

Materieansammlung	-	Zufallsvariable
Masse	-	Wahrscheinlichkeit
dreidimensionaler Raum	-	Menge |R der reellen Zahlen
Punkt im Raum	-	Zahl $x \in \mathbb{R}$
Raumbereich V	-	Zahlenbereich bzw. Intervall [a; b]
Massendichte $\rho(x,y,z)$	-	Wahrscheinlichkeitsdichte $f(x)$
$m(V) = \iiint_V \rho(x,y,z)\,dx\,dy\,dz$	-	$P(a \le X \le b) = \int_a^b f(x)\,dx$

Die Analogie geht sogar noch weiter: Die wichtigste statistische Kenngröße für eine Zufallsvariable, ihr Erwartungswert, entspricht dem Schwerpunkt der Materieansammlung.