Statistik

Vorlesung 4 - diskrete Zufallsvariablen und Verteilungen

Prof. Dr. Sandra Eisenreich

15. April 2024

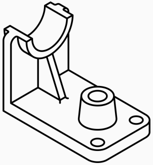
Hochschule Landshut

Agenda

1. diskrete Zufallsvariablen und Verteilungen
2. diskrete Verteilungen

diskrete Zufallsvariablen und Verteilungen

Motivation



Bisher: einzelnes Ereignis

$$A_{10} = \text{“Von 1000 Bauteilen sind 10 niO”}.$$

Wahrscheinlichkeit von 1,2,3,4,5, ... defekten Bauteilen?

$\Rightarrow$ Zufallsvariable:

$$X = \text{“Die Anzahl von niO Bauteilen unter 1000”}$$

A_{10} entspricht dem Ereignis $X = 10$. Ein “Ausgang” von X bei einem Zufallsexperiment heißt Realisierung von X.

- diskreten Zufallsvariable: z.B. ganze Zahlen als Werte;
- stetigen Zufallsvariable: kontinuierliche reelle Zahlen als Werte

Unterschied zwischen Zufallsvariable und Ereignis

Ereignis A:

- wenn ich ein Zufallsexperiment durchführe, trifft ein Ereignis A ein oder nicht.
- z.B. $A = \{(3,5),(4,2),(6,6)\}$

Zufallsvariable X:

- wenn ich ein Zufallsexperiment durchführe, kommt für X eine konkrete Zahl heraus;
- z.B. $X =$ "Summe bei 2x Würfeln"; Zufallsexperiment $\rightarrow$ Ergebnis $(5,3)$ $\rightarrow$ einsetzen $X(5,3) = 8$

Eine Zufallsvariable $X(\omega)$ ist eine Funktion im Ergebnis ω des Zufallsexperiments.

Zufallsvariablen

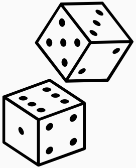
- Ist Ω ein Grundraum, so heißt jede Abbildung

  $$X : \Omega \to \mathbb{R}$$

  von Ω in die Menge $\mathbb{R}$ der reellen Zahlen eine Zufallsvariable auf Ω. $X(\omega)$ heißt auch Realisierung der Zufallsvariable zum Ausgang ω
- Ist der Wertebereich von X abzählbar (z.B. nur die natürlichen Zahlen), heißt X diskrete Zufallsvariable.

Ist der Wertebereich von X nicht abzählbar und erfüllt X eine weitere Eigenschaft, die wir noch kennen lernen werden, heißt X stetige Zufallsvariable.

Beispiel: 2x Würfeln

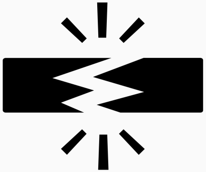


$\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$

- X="die größte Augenzahl beim 2-fachen Würfelwurf",
  - Realisierungen $= \{1,\ldots,6\}$
  - z.B. $X(3,5) = 5$, $X(1,2) = 2$
- X="die Augensumme beim 2-fachen Würfelwurf"
  - Realisierungen $= \{2,\ldots,12\}$
  - z.B. $X(3,5) = 8$, $X(1,2) = 3$

Achtung: Im Allgemeinen können wir aus der Realisierung von X (z.B. 8) nicht das Ergebnis ((3,5)) rekonstruieren (könnte auch (6,2) sein).

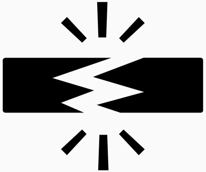
Beispiel: Behebungskosten



Es werden n gleiche Produkte einer Qualitätsprüfung unterzogen. Jedes fehlerhafte Produkt verursacht Behebungskosten in Höhe von K Euro. Bei einem fehlerfreien Produkt fallen keine weiteren Kosten an.

Aufgabe: Beschreiben Sie die insgesamt anfallenden Behebungskosten als Zufallsvariable auf einem geeigneten Grundraum.

Antwort: Behebungskosten



- $\Omega := \{(a_1, \dots, a_n) : a_j \in \{0, 1\} \text{ für } j = 1, \dots, n\}$
- $X(\omega) := K \cdot \sum_{j=1}^{n} a_j$, $\omega = (a_1, \dots, a_n)$ [Gesamt-Behebungskosten]

Notation von Ereignissen mit Zufallsvariablen

Wir schreiben $\{X = k\} := \{\omega \in \Omega : X(\omega) = k\}$ für das Ergebnis, dass X den Wert k annimmt und lassen bei der Wahrscheinlichkeit die Klammern weg, also $P(X = k) := P(\{X = a\})$.

Beispiele:

- Augensumme ist mindestens 10: $\{X \geq 10\} = \{X = 10\} + \{X = 11\} + \{X = 12\}$
- Augensumme liegt zwischen 3 und 8: $\{3 \leq X \leq 8\} = \sum_{k=3}^{8} \{X = k\}$
- Augensumme ist kleiner als 7: $\{X < 7\} = \sum_{k=2}^{6} \{X = k\}$

Zufallsvektoren - Motivation

In einem Wahrscheinlichkeitsraum Ω können mehrere Merkmale gleichzeitig beobachtet werden.

Beispiel:

- 5x Würfeln; X_i = Augenzahl des i-ten Würfel
- Zufallsvektor: $(X_1, X_2, \ldots, X_5)$, eine Realisierung wäre z.B. $(4, 1, 6, 3, 1)$
- die Zufallsvariable $X = X_1 + X_2 + \cdots + X_5$ bezeichnet die Augensumme der 5 Würfel

Zufallsvektoren

- Bei n Zufallsvariablen $(X_1, X_2, \dots, X_n)$ sprechen wir von einem Zufallsvektor.
- wird eine Experiment durchgeführt, so heißt das Ergebnis

$$(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

eine Realisierung des Zufallsvektors.

Unabhängige diskrete Zufallsvariablen

Definition

Zwei diskrete Zufallsvariablen X und Y mit den Wertemengen $W(X)$ und $W(Y)$ heißen unabhängig, wenn für alle möglichen Wertepaare $(x, y) \in (W(X), W(Y))$ gilt:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Dies bedeutet, dass die Ereignisse $\{X = x\}$ und $\{Y = y\}$ für alle Werte x und y unabhängig sind.

Unabhängige diskrete Zufallsvariablen

Der Begriff lässt sich verallgemeinern zu mehreren Zufallsvariablen:

Definition

Die n diskreten Zufallsvariablen $X_1, \ldots, X_n$ heißen unabhängig, wenn für alle möglichen n-Tupel $(x_1, \ldots, x_n)$ aus $W(X_1) \times W(X_2) \times \cdots \times W(X_n)$ gilt:

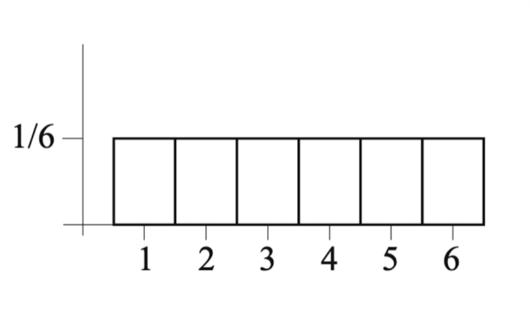
$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$$

diskrete Verteilungen

Motivation: Wahrscheinlichkeitsverteilung einer Zufallsvariable

Ist X eine Zufallsvariable mit Wertemenge $W(X)$, so können wir $P(X = k)$ als Funktion in $k \in W(X)$ sehen. Diese Funktion nennen wir die Verteilung (oder Wahrscheinlichkeitsverteilung/Distribution) der Zufallsvariablen X.

Beispiel: $X =$"Ergebnis bei 1x Würfeln":



Definition (Verteilung)

Es sei X eine diskrete Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, P) und $W = X(\Omega)$ die Wertemenge von X. Dann heißt die Funktion

$$V : W \to [0,1], \quad x \mapsto P(\{\omega \in \Omega : X(\omega) = x\})$$

Verteilung der Zufallsvariablen X. Notation: $V(x) = P(X = x)$

Motivation: Histogramme

Frage: Wie können wir die Verteilung einer Zufallsvariablen plotten?

Idee: zeichne für jeden möglichen Wert ein Säule, die so hoch ist wie ihre Eintrittswahrscheinlichkeit.

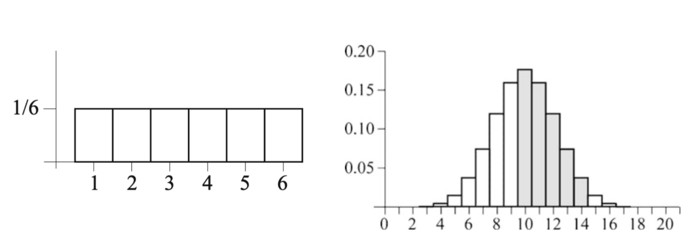
Problem: Wenn es aber sehr viele Werte sind, wird das unübersichtlich; und was wenn man nur die Wahrscheinlichkeit von Intervallen angeben kann?

Lösung: "Binning": teile den Wertebereich in gleich große Intervalle ein und plotte die durchschnittliche Wahrscheinlichkeit jedes Intervalles;

Histogramme

Anleitung zum Zeichnen eines Histogramms:

- teile den Wertebereich in gleich große Intervalle ein;
- berechne die "Durchschnittswahrscheinlichkeit" $\frac{\text{Intervall-Wahrscheinlichkeit}}{\text{Breite des Intervalls}}$ ($\Rightarrow$ Fläche $=$ Intervall-Wahrscheinlichkeit)
- zeichne einen Balken mit dieser Höhe für das Intervall, so dass die Balken aneinanderstoßen



Literatur

- Hartmann, Peter; Mathematik für Informatiker, Springer-Vieweg; 7. Auflage; 2019
- Henze, Norbert; Stochastik für Einsteiger; Springer; 10. Auflage; 2013