

$\Omega = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \{1, 2, 3\}, x_i \neq x_j \text{ für } i \neq j\}$   
 $\# \Omega = 3 \cdot 2 \cdot 1 = 3! = 6$ , Gewichtsfunktion  $p(w) = 1/6, w \in \Omega$   
**Beispiel.** Betrachten wir in Bsp. 1.3, 1 (d) mit  $n = 3$  die Ereignisse

$$A_i = \{(x_1, x_2, x_3) \in \Omega : x_i = i\} \quad \text{für } i = 1, 2, 3$$

Wenn wir eine uniform verteilte Permutation von 1, 2, 3 betrachten, so ist  $A_i = \{i \text{ ist ein Fixpunkt}\}$ . Dann ist die Wahrscheinlichkeit, dass es mindestens einen Fixpunkt gibt:

$$P(A_1 \cap A_2) = \frac{1}{6}$$

$= \{(1, 2, 3)\}$

$$P(A_i) = \frac{\# A_i}{6} = \frac{2}{6} = \frac{1}{3}$$

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) = 1/6 \\ &= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3} \end{aligned}$$

$A_1 = \{(1, 2, 3), (1, 3, 2)\}$

**Bemerkung 1.7.** Wir lassen hier die „philosophische“ Frage offen, welche Interpretation die Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A$  haben soll. Denkbar:

- „naive“ Interpretation der Wahrscheinlichkeit: Die „Natur“ enthält inhärente Unbestimmtheit („ist selbst nicht sicher, was sie tut“), und  $P(A)$  beschreibt den Grad der Sicherheit, mit dem sie sich für das Ereignis  $A$  entscheidet.
- „frequentistische“ Interpretation der Wahrscheinlichkeit: Wenn man das zufällige Experiment unter exakt denselben Bedingungen sehr oft wiederholte, wäre der relative Anteil der Ausgänge, in denen  $A$  eingetreten ist, etwa  $P(A)$ .
- „subjektive“ Interpretation der Wahrscheinlichkeit:  $P(A)$  misst, wie sicher ich mir persönlich bin, dass  $A$  eintreten wird.  
(Beispielsweise: Wieviel Wetteinsatz wäre ich bereit zu bezahlen, wenn mir 1 € ausgezahlt würde, sofern  $A$  eintritt?)

## **Bemerkung 1.7** (Fortsetzung).

Kersting & Wakolbinger [KW, S. vi] schreiben dazu:

„Es kann nicht darum gehen, eine spezielle Intuition gegenüber den anderen durchzusetzen. Dass sich dies innerhalb der Mathematik auch gar nicht als nötig erweist, ist eine der Stärken mathematischer Wissenschaft.“

Siehe z.B. auch die Diskussion in Georgii [Ge, S. 14] am Ende von Abschn. 1.1.3.

## 1.1.1 Zufallsvariablen

### Zufallsvariablen

Oft stellt man sich zumindest intuitiv Zufallsvorgänge so vor, dass ein „zufälliger Wert“ beobachtet oder gemessen wurde, mit dem Verständnis, dass sich bei Wiederholung des Experiments möglicherweise ein anderer Wert aus einer gewissen Menge möglicher Werte ergäbe.

So kann man die Beispiele 1.3 auch folgendermaßen aussprechen:

- 1.(a) Sei  $W$  das Ergebnis eines Wurfs eines fairen 6er-Würfels (Wertebereich  $\{1, 2, 3, 4, 5, 6\}$ ).
- (b) Werfe Würfel dreimal, seien  $W_1, W_2, W_3$  Ergebnisse des 1., 2., 3. Wurfs des Würfels  
( $W = (W_1, W_2, W_3)$  hat Wertebereich  $\{1, 2, 3, 4, 5, 6\}^3$ ).
- (c)  $Z = (X, Y)$  ein zufällig gewähltes Pixel aus  $[0, 1]$

So kann man die Beispiele 1.3 auch folgendermaßen aussprechen:

1.(d)  $n$  verschiedene Eingaben in zufälliger Reihenfolge: Sei

$X = (X_1, X_2, \dots, X_n)$  eine zufällige Permutation von  $1, 2, \dots, n$

(Wertebereich

$\{(x_1, x_2, \dots, x_n) : x_1, \dots, x_n \in \{1, 2, \dots, n\} \text{ paarweise verschieden}\}$ )

2. Sei  $M$  das Ergebnis eines (möglicherweise verfälschten) Münzwurfs

(Wertebereich  $\{\text{Kopf}, \text{Zahl}\}$ )

3. Wähle uniform eine Email aus meiner Inbox (die deutsche und englische Emails enthält, die jeweils Spam sein können oder nicht), sei  $X$  die Sprache und  $Y$  der Spam-Status der betrachteten Email (( $X, Y$ ) hat Wertebereich  $\{\text{Deutsch}, \text{Englisch}\} \times \{\text{Spam}, \text{keinSpam}\}$ )

4. Wir werfen eine faire Münze, bis zum ersten Mal Kopf kommt.

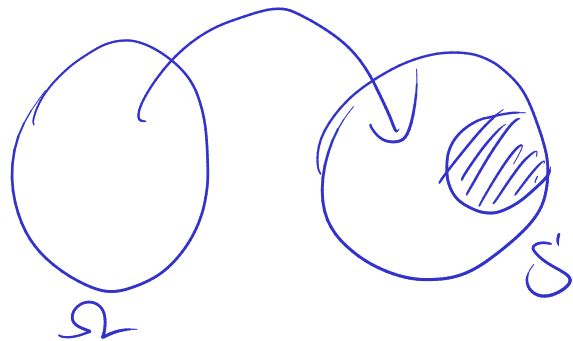
$G$  zählt, wie oft bis dahin Zahl gefallen ist.

**Definition 1.8.**  $(\Omega, \mathcal{F}, P)$  ein Wahrscheinlichkeitsraum,  $S$  eine Menge, eine Abbildung\*  $X : \Omega \rightarrow S$  nennen wir eine *Zufallsvariable*, oft abgekürzt ZV (auch: *Zufallsgröße*) mit Wertebereich  $S$ .

ZVn und Ereignisse:

Für  $A \subset S$  schreibt man

$$\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A)$$



für das Ereignis „ $X$  nimmt einen Wert in  $A$  an“,  $(\Omega, \Sigma)$

falls  $A = \{x\}$ , so schreibt man oft  
 $\{X = x\} := \{X \in \{x\}\},$

falls  $S \subset \mathbb{R}$  auch  $\{X \leq x\} := \{X \in (-\infty, x]\},$   
etc.

\*Im Fall, dass  $S$  überabzählbar ist, muss man strenggenommen *Messbarkeit* fordern, siehe Notizen, Abschnitt 1.2.1.

**Beispiel 1.9** (Indikatorvariable).  $A \in \mathcal{F}$ ,  $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ ,

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A \end{cases}$$

$\mathbf{1}_A$  heißt die Indikatorvariable des Ereignisses  $A$ .

Offenbar  $A = \{ \mathbf{1}_A = 1 \}.$

## Verteilung

**Definition 1.10.** Für eine Zufallsvariable  $X$  mit Wertebereich  $S$  heißt

$$\rho_X : A \mapsto P(X \in A), \quad A \subset S$$

die *Verteilung* von  $X$ .

Eine ZV  $X$  heißt *diskret*, wenn ihr Wertebereich  $S$  (endlich oder) abzählbar ist oder zumindest eine (endliche oder) abzählbare Teilmenge  $S$  enthält mit  $P(X \in S) = 1$ . Die Zahlen

$$\rho_X(\{a\}) := P(X = a), \quad a \in S$$

heißen dann die *Verteilungsgewichte* von  $X$  (oft auch nur: die *Gewichte*).

Abkürzung (mit kleinen „Notationsmissbrauch“):  $\rho_X(a) := \rho_X(\{a\})$ .

## Verteilung (Fall mit Gewichten)

$$\rho_X(A) = P(X \in A), \quad A \subset S, \quad \rho_X(a) = P(X = a), \quad a \in S$$

Offenbar gilt (mit Eigenschaft (A) aus Forderung (1.2)):  $\rho_X(a) \geq 0$ ,

$$\sum_{a \in A} \rho_X(a) = P(X \in A) \quad \text{für } A \subset S, \quad \text{insbes.} \quad \sum_{a \in S} \rho_X(a) = 1$$

und  $\rho_X$  ist ein *Wahrscheinlichkeitsmaß* (auch *Wahrscheinlichkeitsverteilung*) auf  $S$

(d.h.  $\rho_X : \{\text{Teilmengen von } S\} \rightarrow [0, 1]$  und die (analogen) Eigenschaften aus Forderung (1.2) und aus Lemma 1.5 gelten, mit Lesung  $\Omega = S$ ).

Sei  $A = \{a_1, a_2, \dots\}$ , dann sind  $\{X = a_i\} \cap \{X = a_j\} = \emptyset$   
für  $i \neq j$

Man schreibt für die Verteilung einer ZV  $X$  oft auch  $\mathcal{L}(X)$

(das  $\mathcal{L}$  erinnert an English “law” bzw. Französisch «loi»,  
d.h. „Gesetz“),

man schreibt auch  $X \sim \rho$ , wenn  $\mathcal{L}(X) = \rho$  gilt (für eine gewisse  
Wahrscheinlichkeitsverteilung  $\rho$ ).

**Beobachtung 1.11** (Operationen mit Zufallsvariablen). Man kann mit Zufallsvariablen gewissermaßen operieren und „rechnen“ kann wie mit den Variablen in einer Programmiersprache:

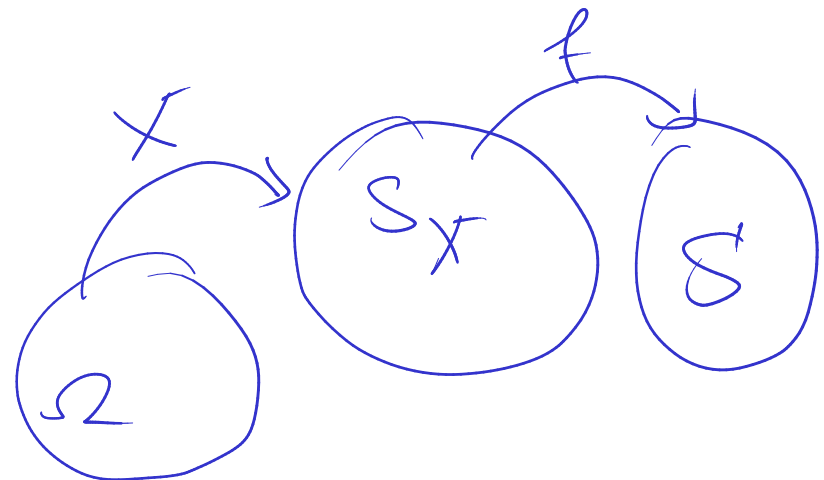
Seien  $X$  und  $Y$  mit Wertebereichen  $S_X$  bzw.  $S_Y$  ZVn auf *demselben Wahrscheinlichkeitsraum*,  $f : S_X \rightarrow S$  eine Abbildung, so können wir die  $S$ -wertige ZV

$$f(X) := f \circ X$$

bilden und die  $S_X \times S_Y$ -wertige ZV  $(X, Y)$ .

Speziell können wir im Fall  $S_X, S_Y \subset \mathbb{R}$  die ZVn  $X^2$ ,  $X + Y$ ,  $X - Y$ ,  $XY$ , etc. bilden.

(d.h.  
 $X: \Omega \rightarrow S_X$   
 $Y: \Omega \rightarrow S_Y$ )



## Beobachtung 1.11 (Operationen mit Zufallsvariablen, 2).

Mit  $f : S_X \rightarrow S$  wie oben ergibt sich (im diskreten Fall) für die Gewichte von  $f(X)$  dann

$$s \in S$$

$$\begin{aligned} p_{f(X)}(s) &= P(f(X)=s) = P(X \in \{x \in S_X : f(x)=s\}) \\ &= \sum_{x: f(x)=s} \underbrace{P(X=x)}_{=p_X(x)} = \sum_{x: f(x)=s} p_X(x) \end{aligned}$$

## Beispiel (Verteilung der Augensumme zweier fairer Würfel).

Sei  $(W_1, W_2)$  das Ergebnis des 1. und 2. Wurf eines fairen Würfels,  
 $f : \{1, 2, \dots, 6\}^2 \rightarrow \{2, 3, \dots, 12\}$ ,  $(w_1, w_2) \mapsto w_1 + w_2$ , so ist

$Y := W_1 + W_2$  hat Wertebereich  $S_Y = \{2, 3, \dots, 12\}$

$$\begin{aligned} P(W_1 + W_2 = s) &= P(f(W_1, W_2) = s) = P((W_1, W_2) \in \\ &= \sum_{(w_1, w_2) : w_1 + w_2 = s} \underbrace{P((W_1, W_2) = (w_1, w_2))}_{= \frac{1}{36}} \end{aligned}$$

$\{(w_1, w_2) \in \{1, \dots, 6\}^2 : w_1 + w_2 = s\}$

NR

$$2 = 1 + 1$$

$$3 = 1 + 2 = 2 + 1$$

$$4 = 1 + 3 = 2 + 2 \\ = 3 + 1$$

$$\begin{aligned} &= \frac{1}{36} \# \{(w_1, w_2) \in \{1, \dots, 6\}^2 : w_1 + w_2 = s\} \\ &= \frac{6 - |7 - s|}{36}, \quad s \in S_Y \end{aligned}$$

## Hinweis

### **Bemerkung 1.12** (Kanonisches Modell für eine ZV).

Man kann eine Zufallsvariable  $X$  mit Wertebereich  $S$  und Verteilung  $\rho$  stets in „kanonischer Weise“ mit der Wahl  $\Omega = S$  und  $P = \rho$  und geeignetem  $\mathcal{F} \subset 2^S$  (im diskreten Fall kann man  $\mathcal{F} = 2^S$  wählen) als  $X = \text{Id}_S$  auf dem  $W$ -raum  $(S, \mathcal{F}, \rho)$  formulieren.

Man kann daher genausogut die mathematische Modellierung eines Zufallsphänomens mit der Formulierung geeigneter Zufallsvariablen samt Verteilung beginnen, siehe auch Diskussion und Referenzen in Abschnitt 1.2.2 der Notizen.

## 1.1.2 diskrete Verteilungen

### Urnenmodelle

**Beispiel 1.13** (Ziehen mit und ohne Zurücklegen: Urnenmodelle).

Eine Urne enthalte  $n$  (mit  $1, 2, \dots, n$  nummerierte) Kugeln, wir ziehen zufällig  $k$  ( $\leq n$ ) heraus. Sei

$X_i$  die Nummer der Kugel im  $i$ -ten Zug,  $i = 1, \dots, k$

und

$$X = (X_1, X_2, \dots, X_k)$$

$n$  Kugeln in Urne,

$X_i$  die Nummer der Kugel im  $i$ -ten Zug,  $i = 1, \dots, k$

Wir betrachten 4 mögliche Situationen:

1. Ziehen mit Zurücklegen, mit Beachtung der Reihenfolge:

$X = (X_1, \dots, X_k)$  hat Wertebereich

$$W_1 = \{1, 2, \dots, n\}^k (= \{(x_1, \dots, x_k) : x_i \in \{1, \dots, n\}, i = 1, \dots, k\})$$

$$P(X = (x_1, \dots, x_k)) = \frac{1}{\#W_1} = \frac{1}{n^k} \quad \text{für jedes } (x_1, \dots, x_k) \in W_1$$

$n$  Kugeln in Urne,

$X_i$  die Nummer der Kugel im  $i$ -ten Zug,  $i = 1, \dots, k$

2. Ziehen ohne Zurücklegen, mit Beachtung der Reihenfolge:

$X$  hat Wertebereich

$$W_2 = \{(x_1, \dots, x_k) : x_1, \dots, x_k \in \{1, \dots, n\}, x_i \neq x_j \text{ für } i \neq j\}$$

$$P(X = (x_1, \dots, x_k)) = \frac{1}{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)} = \frac{(n-k)!}{n!}$$

$$\left( = \frac{1}{\#W_2} \right) \text{ für jedes } (x_1, \dots, x_k) \in W_2$$

3. Ziehen ohne Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge:  
Wir beobachten nicht  $X$  (das hier wie in 2. verteilt wäre), sondern mit

$$\varphi((x_1, \dots, x_k)) = \{x_1, \dots, x_k\} \quad \begin{array}{l} \text{(verwandle Vektor in Menge,} \\ \text{d.h. vergiss die Reihenfolge)} \end{array}$$

(nur)  $Y = \varphi(X)$  mit Werten in  $W_3 = \{A \subset \{1, 2, \dots, n\} : \#A = k\}$ .

Für  $A \in W_3$ :

$$\begin{aligned} P(Y=A) &= P(\varphi(X)=A) = P(X \in \varphi^{-1}(A)) \\ &= \sum_{x \in W_2 : \varphi(x)=A} P(X=x) = \underbrace{k! \cdot \frac{(n-k)!}{n!}}_{=\frac{(n-k)!}{n!}} = \frac{1}{\binom{n}{k}}, \end{aligned}$$

d.h.  $Y$  ist uniform auf  $W_3$

## 4. Ziehen mit Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge:

Sei  $X = (X_1, \dots, X_k)$  wie in 1., wir beobachten aber (nur)  $Y = (Y_1, \dots, Y_n)$ , wobei

$$Y_j = \#\{1 \leq i \leq k : X_i = j\} \quad \text{für } j = 1, \dots, n$$

( $Y_j$  gibt an, wie oft Kugel  $j$  gezogen wurde.)

$Y$  hat Wertebereich  $\{(l_1, l_2, \dots, l_n) \in \mathbb{N}_0^n : l_1 + l_2 + \dots + l_n = k\}$

Für  $(l_1, \dots, l_n) \in W_k$  gibt es

$$\frac{k!}{l_1! \cdot l_2! \cdot \dots \cdot l_n!} =: \binom{k}{l_1 \ l_2 \ \dots \ l_n}$$

(„Multinomial-  
koeffizient“)

Wahlen von  
 $(X_1, \dots, X_k) \in W_k$

mit  $\#\{1 \leq i \leq k : X_i = j\} = l_j, j = 1, \dots, n$

Zähler:

Platze  $l_1$  1-er:  $\frac{k(k-1)\dots(k-l_1-1)}{l_1!}$ , platze 2-er, ...

## 4. Ziehen mit Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge:

Sei  $X = (X_1, \dots, X_k)$  wie in 1., wir beobachten aber (nur)  $Y = (Y_1, \dots, Y_n)$ , wobei

$$Y_j = \#\{1 \leq i \leq k : X_i = j\} \quad \text{für } j = 1, \dots, n$$

( $Y_j$  gibt an, wie oft Kugel  $j$  gezogen wurde.)

somit

$$P(Y = (l_1, \dots, l_n)) = \binom{k}{l_1 \ l_2 \ \dots \ l_n} \cdot \frac{1}{n^k}$$

für  $(l_1, \dots, l_n) \in W_4$

(eine  
„Multinomial-  
verteilung“)

## Bemerkung 1.14. Es gilt

$$|W_4| = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

(Insbesondere ist die Verteilung auf  $W_4$  aus Beispiel 1.13, 4. nicht die uniforme.)

Die uniforme Verteilung auf dem  $W_4$  aus Beispiel 1.13, 4. heißt auch die „Bose-Einstein-Verteilung“, die in Beispiel 1.13, 4. betrachtete Verteilung heißt die „Maxwell-Boltzmann-Verteilung“.

**Beispiel.** Eine Hörsaalreihe habe  $n$  Plätze, darauf nehmen  $m$  ( $\leq n/2$ ) Männer und  $n - m$  Frauen rein zufällig Platz.

Die Wahrscheinlichkeit, dass (in diesem Modell) keine zwei Männer nebeneinander sitzen

**Beispiel 1.15** (Hypergeometrische Verteilung). Eine Urne enthalte  $n$  Kugeln, davon  $s$  schwarze und  $w$  weiße ( $s + w = n$ ), ziehe  $k$ -mal ohne Zurücklegen,

$$\text{Hyp}_{s,w,k}(\{\ell\}) = \frac{\binom{s}{\ell} \binom{w}{k-\ell}}{\binom{s+w}{k}}, \quad \ell = 0, 1, \dots, k$$

ist die W'keit, genau  $\ell$  schwarze Kugeln zu ziehen.

## Beispiel 1.16 ( $p$ -Münzwurf und Binomialverteilung).

1.  $S = \{0, 1\}$ ,  $\text{Ber}_p(\{1\}) = p = 1 - \text{Ber}_p(\{0\})$  mit einem  $p \in [0, 1]$   
(„Bernoulli-Verteilung“<sup>1</sup>)

2.  $n$ -facher  $p$ -Münzwurf (mit  $p \in [0, 1]$ ):  $S = \{0, 1\}^n$ ,

$$\text{Ber}_p^{\otimes n}(\{(x_1, \dots, x_n)\}) = p^{|\{i \leq n : x_i = 1\}|} (1 - p)^{|\{i \leq n : x_i = 0\}|}$$

Für  $X = (X_1, \dots, X_n) \sim \text{Ber}_p^{\otimes n}$  sagt man:  $X_1, \dots, X_n$  unabhängig und  $X_i \sim \text{Ber}_p$  für  $i = 1, 2, \dots, n$

3. Binomialverteilung (zum Parameter  $n$  und  $p$ ,  $n \in \mathbb{N}$ ,  $p \in [0, 1]$ ):

$$\text{Bin}_{n,p}(\{k\}) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in S := \{0, 1, \dots, n\}$$

(dies ist die  $W$ 'keit, beim  $n$ -fachen Münzwurf genau  $k$  Erfolge zu beobachten)

<sup>1</sup>nach Jakob Bernoulli, 1654–1705

## Beispiel 1.16, 3. mit Zufallsvariablen ausgesprochen:

$$X = (X_1, \dots, X_n) \sim \text{Ber}_p^{\otimes n}, \quad Y := X_1 + X_2 + \dots + X_n,$$

so ist  $Y \sim \text{Bin}_{n,p}$ , denn für  $k \in \{0, 1, \dots, n\}$  und für jede Wahl von  $(x_1, \dots, x_n) \in \{0, 1\}^n$  mit  $x_1 + \dots + x_n = k$  ist

$$P(X_1 = x_1, \dots, X_n = x_n) = p^k (1-p)^{n-k}, \text{ somit mit}$$

$$B_k = \{(x_1, \dots, x_n) \in \{0, 1\}^n : x_1 + \dots + x_n = k\}$$

$$\begin{aligned} P(Y = k) &= P((X_1, \dots, X_n) \in B_k) \\ &= \sum_{(x_1, \dots, x_n) \in B_k} P((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{(x_1, \dots, x_n) \in B_k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

$$\text{da } \#B_k = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} = \binom{n}{k}.$$

Insoweit ist dies eine Beispiel-Instanz zu Beobachtung 1.11.