# Vorlesung 2b

# Diskrete Zufallsvariable

# und ihre Verteilungen

mit den Beispielen

Anzahl der Erfolge beim $n$-fachen $p$-**Münzwurf**

und

Besetzungen beim $n$-fachen $(p_1, \ldots, p_g)$-**Würfeln**

Inhalt:

1. Die Grundbegriffe
2. Zufällige Paare und ihre Komponenten
3. Weiterverarbeitung von Zufallsvariablen und Transport von Verteilungen
4. Die Anzahl der Erfolge beim fairen **Münzwurf**
5. Die Anzahl der Sechsen beim fairen **Würfeln**
6. Vom $p$-Münzwurf zur **Binomialverteilung**
7. Vom Ziehen mit Zurücklegen zum $p$-Münzwurf
8. Vom $p$-Münzwurf zum $(p_1, \ldots, p_g)$-Würfeln
9 . Vom $(p_1, \ldots, p_g)$- Würfeln zur **Multinomialverteilung**

# 1. Die Grundbegriffe

(Buch S. 20-21)

Bisher hatten wir uns mit Zufallsvariablen beschäftigt, deren Wertebereich $S$ endlich war.

Die (schon in Vorlesung 1b formulierten)

zwei Grundregeln

für Wahrscheinlichkeiten lauteten für diesen Fall:

Normiertheit auf Eins:

$$\mathbf{P}(X \in S) = 1 \, .$$

Additivität:

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a), \qquad A \subset S$$

Diese beiden Regeln behalten ihren Sinn, wenn der Wertebereich nicht endlich, sondern abzählbar unendlich ist.

Beispiel: $S = \mathbb{N}$

$$\mathbb{P}(X = 1) = \frac{1}{2}, \quad \mathbb{P}(X = 2) = \frac{1}{4}, \quad \mathbb{P}(X = 3) = \frac{1}{8}, \quad \ldots$$

$$\mathbb{P}(X = a) = 1/2^a, \quad a \in \mathbb{N}.$$

Auch wenn der Wertebereich von $X$

eine überabzählbare Menge ist

(wie z.B. die Menge der reellen Zahlen $\mathbb{R}$

oder das “Einheitsintervall” $[0, 1]$

oder das “Einheitsquadrat” $[0, 1] \times [0, 1]$),

behalten beide Regeln ihren Sinn, wenn man fordert,

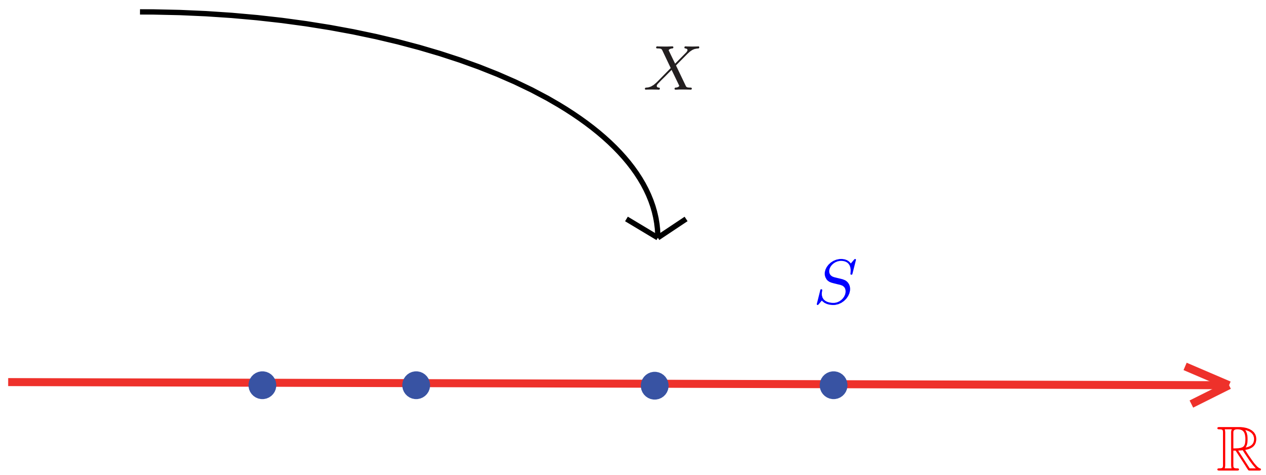
dass der Wertebereich

eine endliche oder abzählbar unendliche Menge $S$ enthält

mit

$$\mathbf{P}(X \in S) = 1.$$

# Beispiel: Wertebereich $\mathbb{R}$




$S \subset \mathbb{R}$ endlich oder abzählbar unendlich mit $\mathbf{P}(X \in S) = 1$

Warum ist das interessant?

Wie wir sehen werden (und wie jetzt schon intuitiv klar ist),
kann man mit reellwertigen Zufallsvariablen rechnen.

Man kann z.B. eine reelwertige Zufallsvariable $X$ halbieren,
und wenn die Zufallsvariable $X$ diskret ist,
ist auch die Zufallvariable $X/2$ diskret.

Definition:

Eine Zufallsvariable $X$ heißt **diskret**,

falls ihr Wertebereich

eine diskrete (d.h. endliche oder abzählbar unendliche)

Menge $S$ enthält mit

$$\mathbf{P}(X \in S) = 1.$$

Für diskrete Zufallsvariable $X$ und

$$\mathbf{P}(X \in S) = 1$$

mit einer endlichen oder abzählbar unendlichen Menge $S$ gilt:

$$\mathbf{P}(X \in A) = \sum_{a \in A} \mathbf{P}(X = a), \qquad A \subset S$$

(Additivität)

Die Zahlen $\rho(a) := \mathbf{P}(X = a), \qquad a \in S$ ,

sind die **Verteilungsgewichte**.

Die Abbildung $A \mapsto \rho(A) := \mathbf{P}(X \in A), \qquad A \subset S$,

heißt die **Verteilung** von $X$.

# 2. Zufällige Paare und ihre Komponenten

(Buch S. 21)

$X = (X_1, X_2)$ sei eine Zufallsvariable

mit $\mathbf{P}(X_i \in S_i) = 1,\ \ S_i$ diskret, $i = 1, 2$.

Dann ist auch $X$ diskret, mit $\mathbf{P}(X \in S_1 \times S_2) = 1$.

Die Ereignisse $\{(X_1, X_2) = (a_1, a_2)\}$ notieren wir auch als

$$\{X_1 = a_1, X_2 = a_2\}.$$

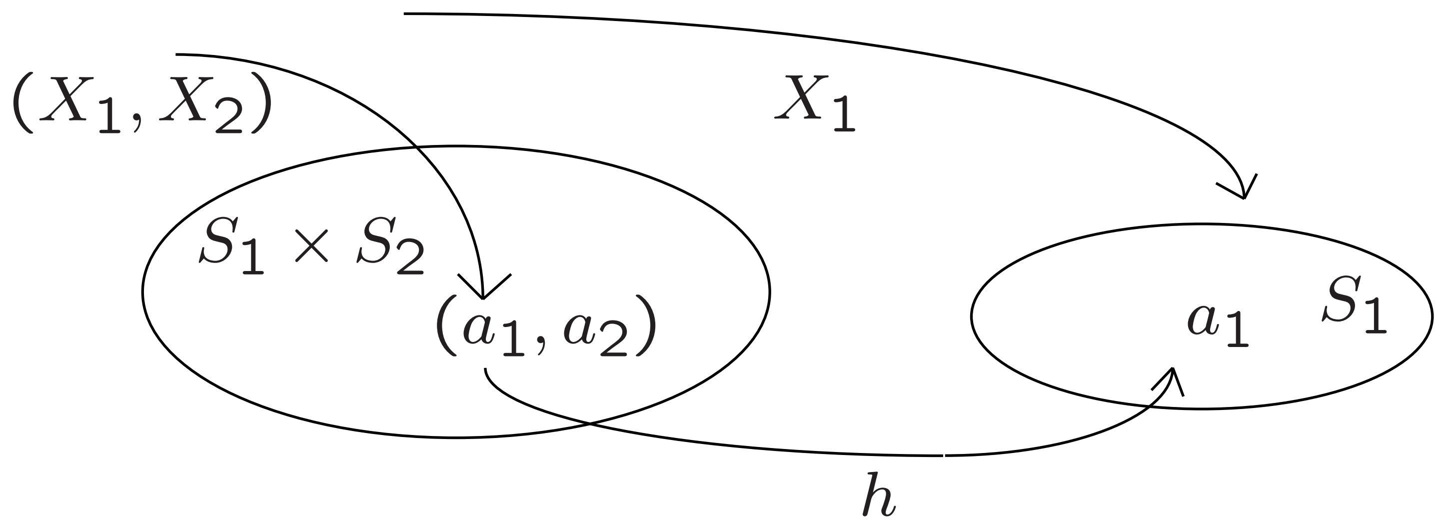
Die Verteilungsgewichte von $X$ schreiben wir als

$$\begin{aligned}\rho(a_1, a_2) &= \mathbf{P}\big((X_1, X_2) = (a_1, a_2)\big)\\ &= \mathbf{P}(X_1 = a_1, X_2 = a_2)\ ,\end{aligned}$$

Sei $\rho_1$ die Verteilung von $X_1$.

Man erhält deren Gewichte als

$$\rho_1(a_1) = \sum_{a_2 \in S_2} \rho(a_1, a_2)\ .$$




$$h((a_1, a_2)) := a_1$$

ist die

Projektion des Paares $(a_1, a_2)$ auf seine erste Komponente

# 3. Weiterverarbeitung von Zufallsvariablen

# und

# Transport von Verteilungen

(Buch S. 21-22)

Der Übergang von $X = (X_1, X_2)$

zur Komponente $X_1$

ist ein Beispiel einer

*Vergröberung* (*Weiterverarbeitung*) einer Zufallsvariablen:

$$X_1 = h(X)$$

mit $h((a_1, a_2)) := a_1$

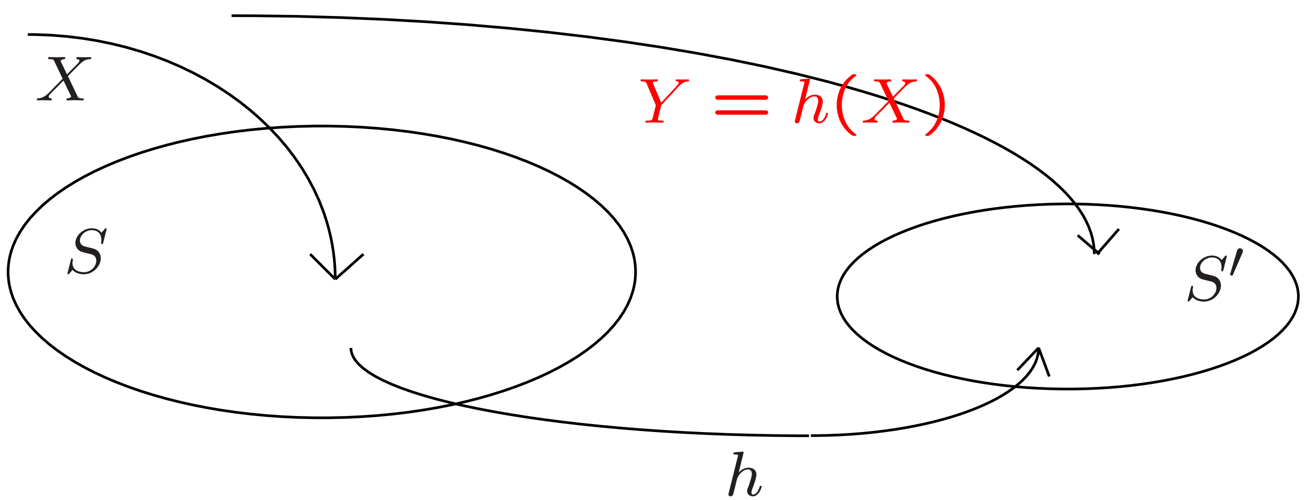
Sind $S$ und $S'$ zwei Mengen,

$X$ eine Zufallsvariable mit Zielbereich $S$,

$h$ eine Abbildung von $S$ nach $S'$,

und nimmt man $X$ als zufällige Eingabe von $h$,

dann bekommt man eine Zufallsvariable $Y$ mit Zielbereich $S'$:

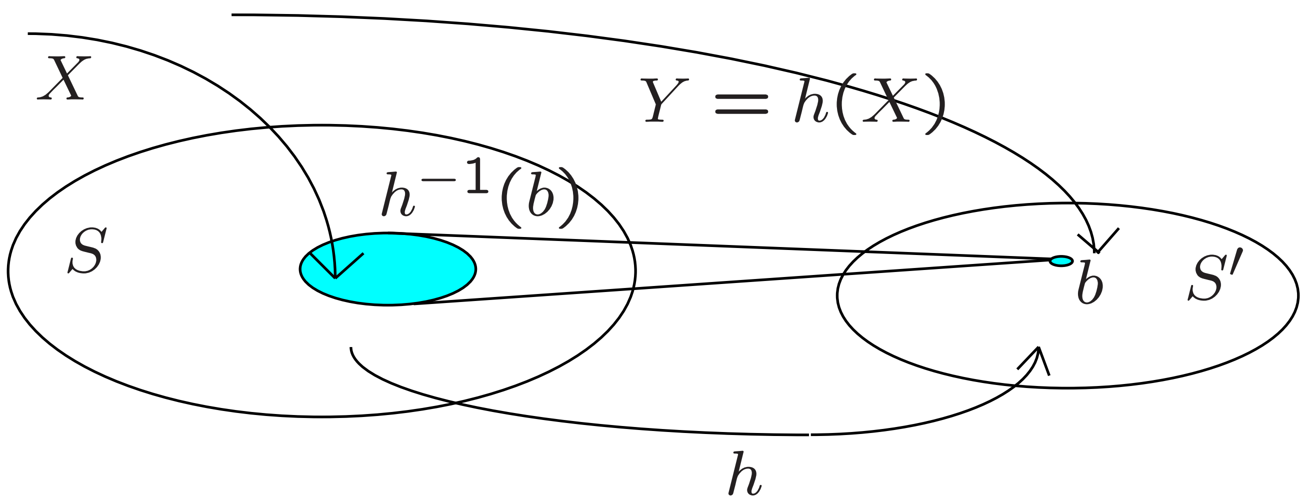


Für jedes $b \in S'$ gilt:

$$\{h(X) = b\} = \{X \in h^{-1}(b)\}$$

Für die Verteilungsgewichte von $Y = h(X)$ ergibt sich:

$$\mathbf{P}(Y = b) = \mathbf{P}(X \in h^{-1}(b)) = \sum_{a \in h^{-1}(b)} \mathbf{P}(X = a).$$

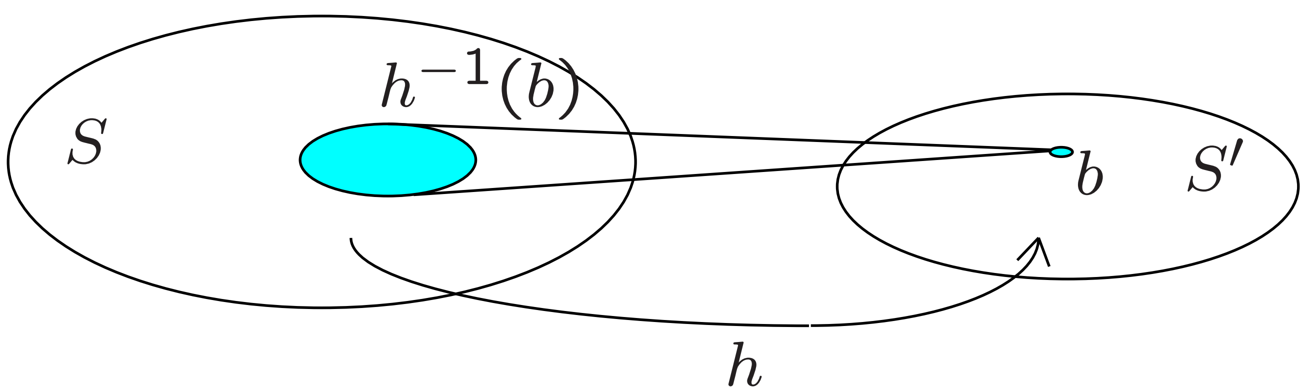


Bezeichnet $\rho$ die Verteilung von $X$ und $\rho'$ die von $Y$,

dann ist

$$\rho'(b) = \sum_{a \in h^{-1}(b)} \rho(a).$$

Man sagt: Die Verteilung $\rho$ wird durch die Abbildung $h$ in die Verteilung $\rho'$ transportiert.



Diese Situation haben wir schon mehrmals angetroffen:

in Vorlesung 1b:

$X :=$ rein zufällige $1, \ldots, g$-Folge der Länge $n$

$T = h(X)$:= Zeitpunkt der ersten Kollision

(mit $T := \infty$ falls keine Kollision eintritt)

in Vorlesung 2a:

$X :=$ rein zufällige Permutation von $1, \ldots, n$

$h(X) :=$ Länge des Zyklus von $X$, der die Eins enthält.

Heutiges Programm:

Weitere Beispiele für

“Vergröberungen von zufälligen Folgen”

$\rightarrow$ wichtige Beispiele

diskreter Zufallsvariabler und diskreter Verteilungen.

# 4. Die Anzahl der Erfolge beim fairen Münzwurf

(vgl. Buch S. 22)

$$S := \{0,1\}^n$$

die Menge der 01-Folgen der Länge $n$

$X$ sei uniform verteilt auf $S$,

jeder Ausgang hat somit das Gewicht

$$\frac{1}{2^n} = \frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}$$

Man sagt dann auch:

$X$ ist ein $n$-facher “fairer Münzwurf”.

$Y :=$ die Anzahl der Einsen in $X$.

Was ergibt sich für die Verteilungsgewichte von $Y$?

Jede einzelne 01-Folge $a$ der Länge $n$ mit genau $k$ Einsen

hat Gewicht

$$\frac{1}{2^n}$$

Wieviele derartige $a$ gibt es?

$$\binom{n}{k}$$

Also:

$$\mathbf{P}(Y = k) = \binom{n}{k}\frac{1}{2^n}, \quad k = 0, \ldots, n.$$

# 5. Die Anzahl der Sechsen beim fairen Würfeln

(vgl. S. 28)

## Beispiel

$n$-faches Würfeln:

Wie ist die Anzahl der Sechsen verteilt?

$W = (W_1, \dots, W_n)$ uniform verteilt auf

$S := \{1, \dots, 6\}^n$.

$Z := (Z_1, \dots, Z_n)$, mit

$Z_i := \mathbf{1}_{\{6\}}(W_i)$

$Z$ ist also eine zufällige $01$-Folge, mit

$Z_i = 1$ falls der $i$-te Wurf eine Sechs ergibt

und $Z_i = 0$ sonst.

Wie ist $Z$ verteilt?

$$\mathbf{P}(Z_1 = 1, \ldots, Z_k = 1, Z_{k+1} = 0, \ldots, Z_n = 0)$$

$$= \mathbf{P}(W_1 = 6, \ldots, W_k = 6, W_{k+1} \neq 6, \ldots, W_n \neq 6)$$

$$= \frac{1^k \cdot 5^{n-k}}{6^n}$$

$$= p^k q^{n-k},$$

mit $p := \frac{1}{6}$ und $q := \frac{5}{6}$.

Auch für jede andere Platzierung von genau $k$ “Sechsen” in den $n$ Würfen ergibt sich diese W’keit.

Fazit zur

Verteilung der Anzahl der Sechsen beim $n$-fachen Würfeln:

Sei

$W = (W_1, \dots, W_n)$ uniform verteilt auf $S := \{1, \dots, 6\}^n$,

$Z := (Z_1, \dots, Z_n), \quad$ mit $Z_i := \mathbf{1}_{\{6\}}(W_i)$

Die Verteilung von

$$Y := Z_1 + \cdots + Z_n$$

ist dann gegeben durch die Gewichte

$\mathbf{P}(Y = k) = \binom{n}{k} p^k q^{n-k} \quad$ (warum?)

# 6. Vom $p$-Münzwurf zur Binomialverteilung

(“Was $1/6$ recht ist, soll $p$ billig sein...”)

(Buch S. 22)

Definition ($p$-Münzwurf):

Sei $p \in [0,1]$, $q := 1 - p$.

Eine Zufallsvariable $Z$ mit Zielbereich

$$S = \{0,1\}^n = \{a = (a_1, \ldots, a_n) \ : a_i \in \{0,1\}\}$$

heißt $n$-facher $p$-Münzwurf,

wenn für alle $a \in S$ mit $k$ Einsen und $n-k$ Nullen gilt:

$$\mathbf{P}(Z = a) = p^k q^{n-k}.$$

Ein Paradebeispiel für die

Weiterverarbeitung einer Zufallsvariablen ist die

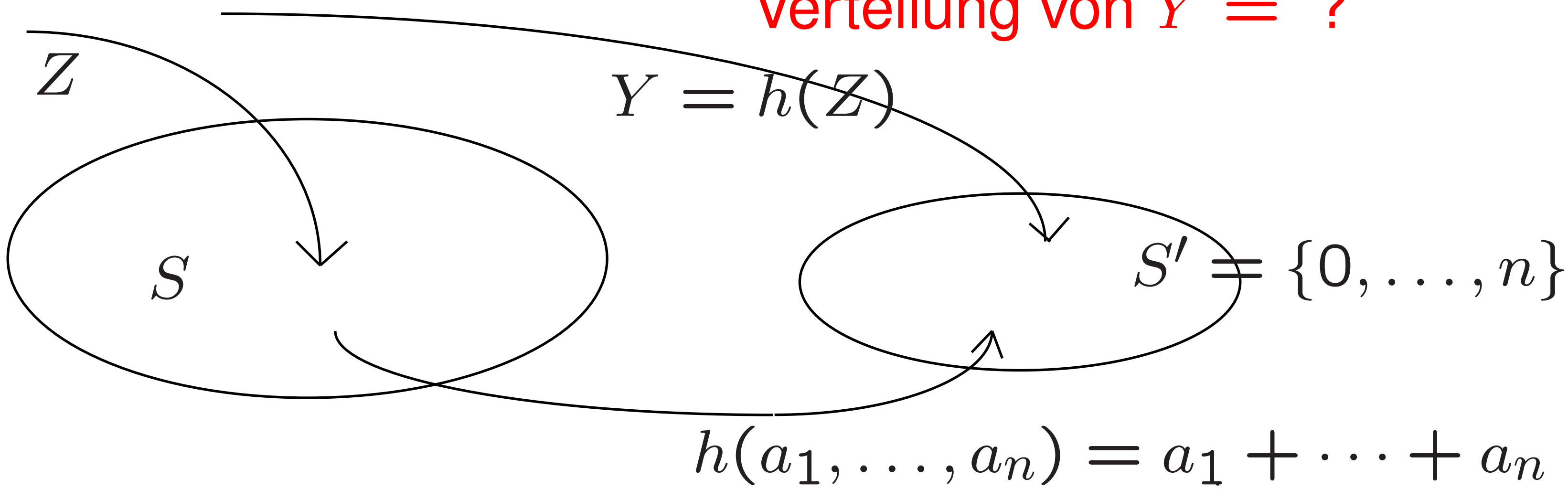
*Anzahl der Erfolge beim $n$-fachen $p$-Münzwurf*:

Sei $Z = (Z_1, \ldots, Z_n)$ ein $n$-facher $p$-Münzwurf

und $Y = Z_1 + \cdots + Z_n =: h(Z)$ die *Anzahl der Erfolge*

(die Anzahl der Einsen in der zufälligen 0-1 Folge $Z$)

Verteilung von $Y =$ ?

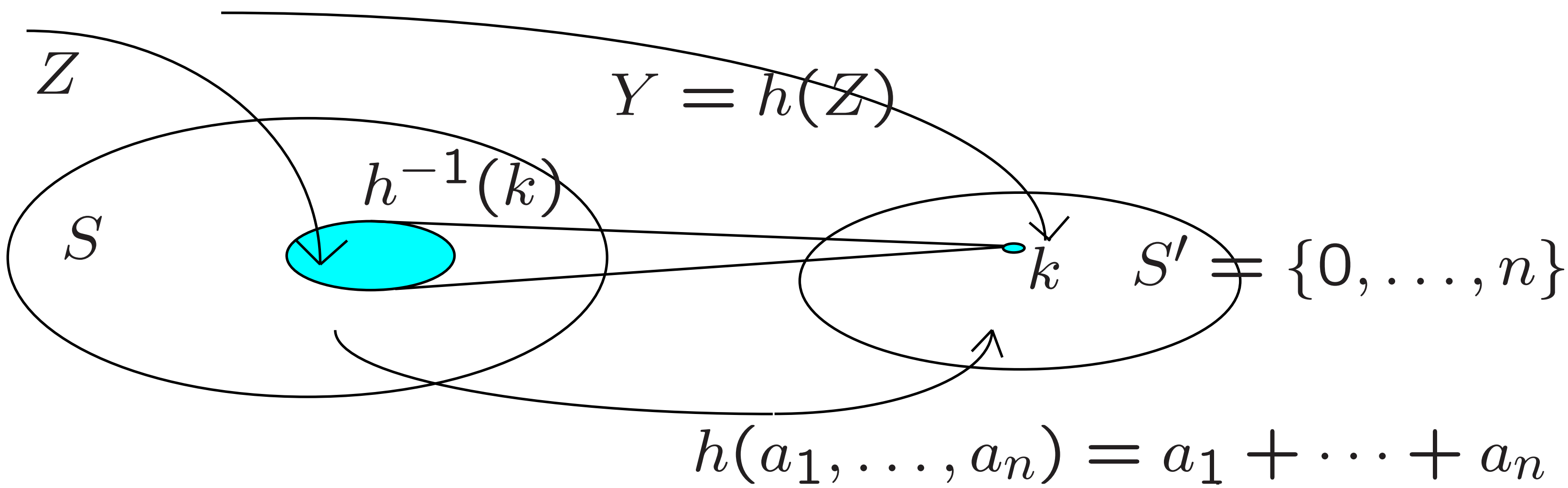


Jedes $a \in S$ mit $h(a) = k$

(d.h. mit $k$ Einsen und $n - k$ Nullen)

hat Gewicht $p^k\,(1-p)^{n-k}$.

Wieviele solche $a$ gibt es?



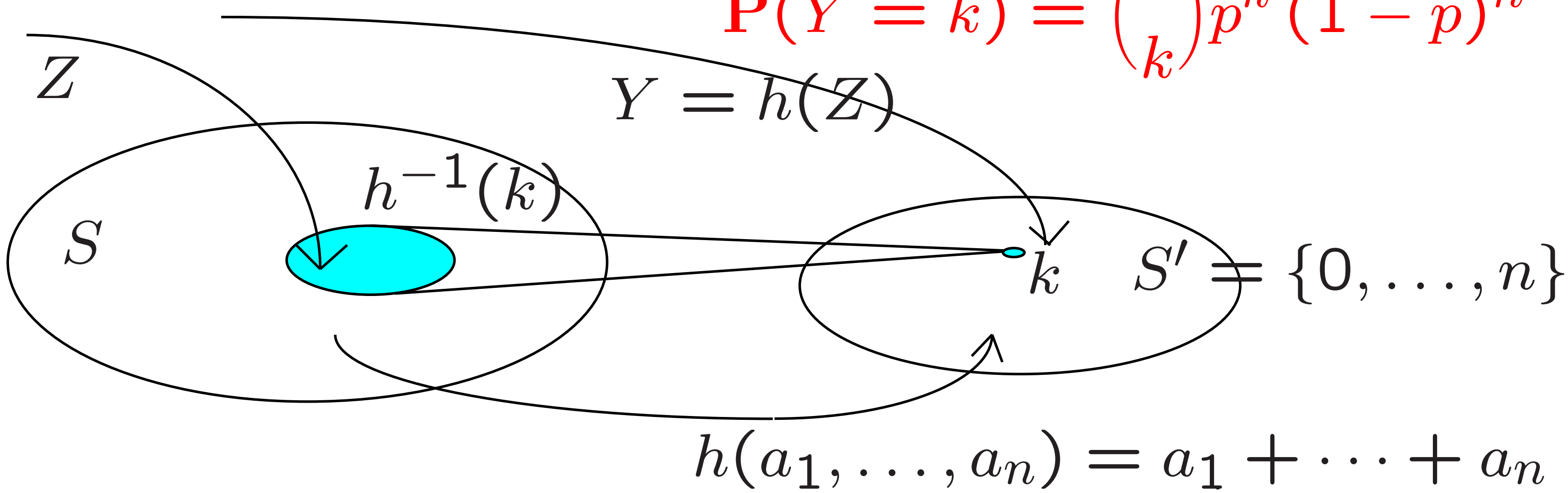
Jedes $a \in S$ mit $h(a) = k$

(d.h. mit $k$ Einsen und $n - k$ Nullen)

hat Gewicht $p^k\,(1-p)^{n-k}$.

Es gibt $\binom{n}{k}$ solche $a$.

$$\mathbf{P}(Y = k) = \binom{n}{k} p^k\,(1-p)^{n-k}$$



Definition:

Eine Zufallsvariable $X$ mit Zielbereich $\{0, 1, \ldots, n\}$

heißt *binomialverteilt* mit Parametern $n$ und $p$,

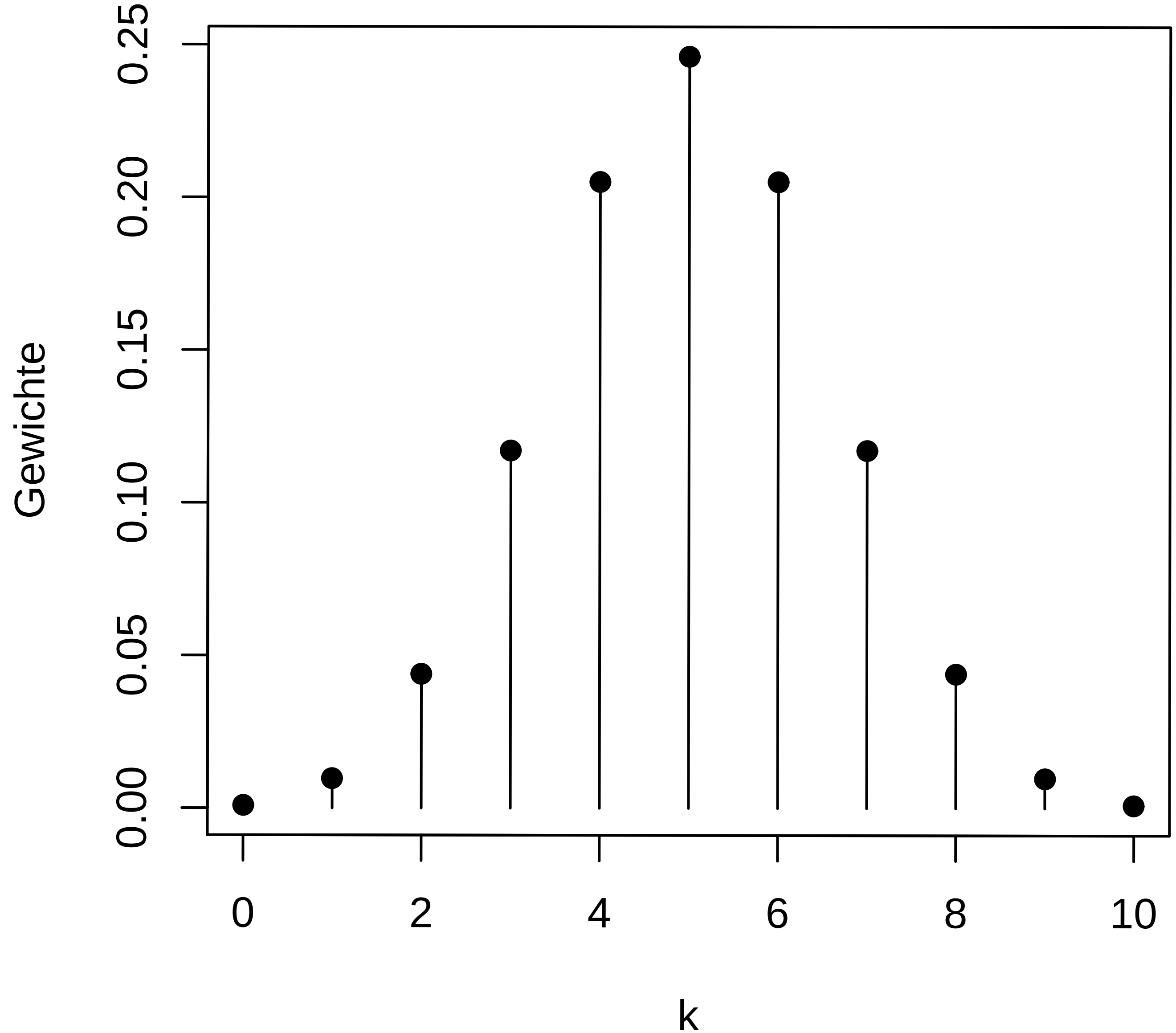
kurz

Bin$(n, p)$-verteilt,

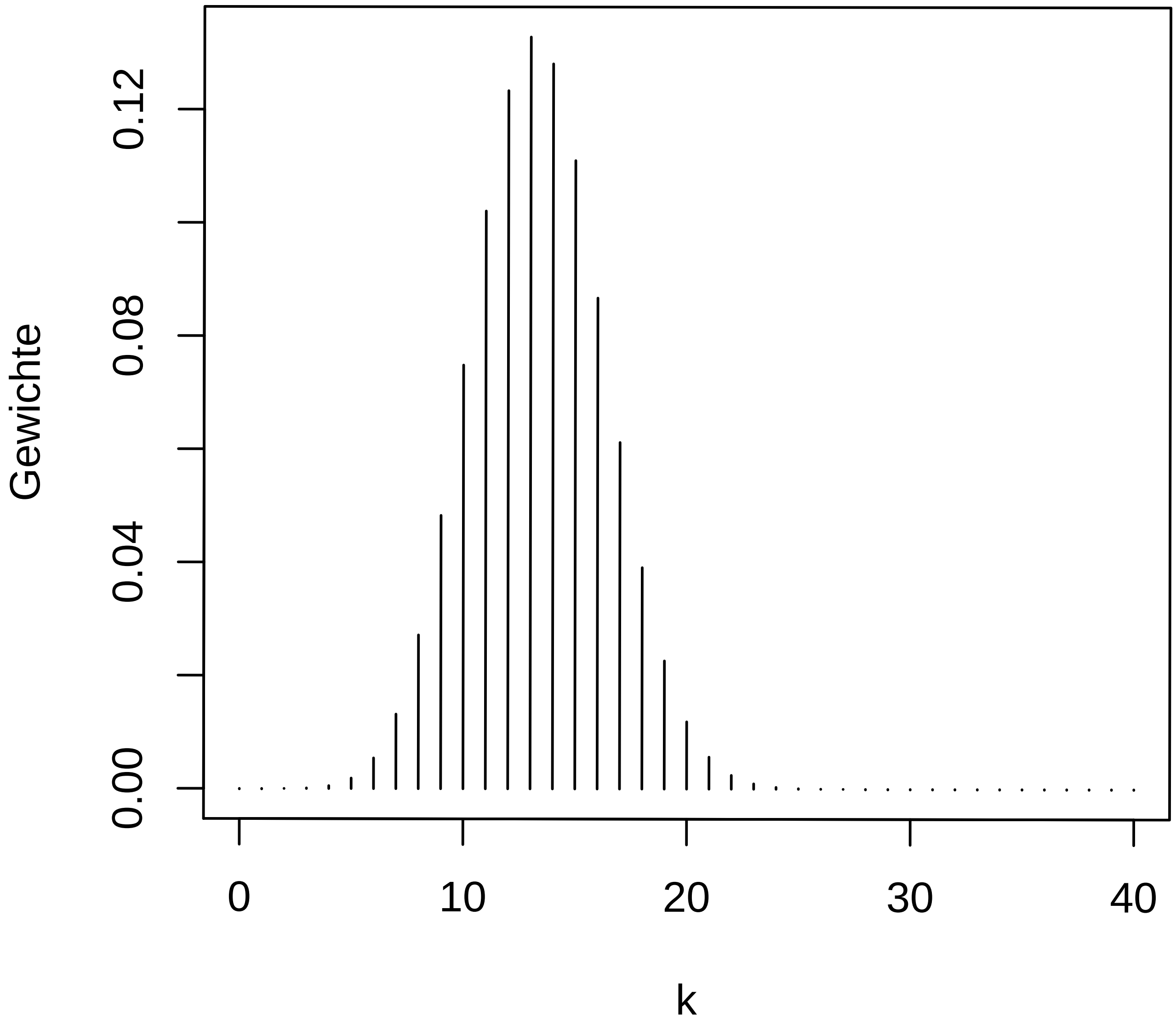
wenn

$$\mathbf{P}(X = k) = \binom{n}{k} p^k q^{n-k} \,, \quad k = 0, 1, \ldots, n \,,$$

mit $q = 1 - p$.




Gewichte der Bin(10, 1/2) Verteilung




Gewichte der Bin(40, 1/3) Verteilung

# 7. Vom Ziehen mit Zurücklegen zum $p$-Münzwurf

(vgl. Buch S. 33)

$n$-maliges Ziehen *mit Zurücklegen*

aus einer ideal durchmischten Urne.

Ein Anteil $p$ der Kugeln ist rot,

der restliche Anteil $q = 1 - p$ ist blau.

Zufällige 0-1 Folge $Z = (Z_1, \ldots, Z_n)$:

$Z_i = 1$ wenn beim $i$-ten Zug eine rote Kugel kommt,

und $Z_i = 0$ wenn beim $i$-ten Zug eine blaue Kugel kommt.

Sei $a$ eine vorgegebene 0-1 Folge der Länge $n$ mit $k$ Einsen,

z. B.: $a := (\underbrace{1,\ldots,1}_{k\text{-mal}},\ \underbrace{0,\ldots,0}_{(n-k)\text{-mal}})$

$$\mathbf{P}(Z = a) = ?$$

Sei $g$ die Gesamtanzahl der Kugeln in der Urne.

Die Anzahl der roten Kogeln ist $pg$,

die der blauen Kugeln ist $qg$. Für obiges $a$ gilt:

$$\mathbf{P}(Z = a) = \frac{(pg)^k (qg)^{n-k}}{g^n} = p^k q^{n-k}$$

Das ist so für jede 0-1 Folge $a$ mit $k$ Einsen und $n - k$ Nullen.

Zur Erinnerung:

Definition ($p$-Münzwurf):

Sei $p \in [0,1]$, $q := 1 - p$.

Eine Zufallsvariable $Z$ mit Zielbereich

$$S = \{0,1\}^n = \{a = (a_1, \ldots, a_n)\ : a_i \in \{0,1\}\}$$

heißt $n$-facher $p$-Münzwurf,

wenn für alle $a \in S$ mit $k$ Einsen und $n-k$ Nullen gilt:

$$\mathbf{P}(Z = a) = p^k q^{n-k}.$$

# 8. Vom $p$-Münzwurf zum $(p_1, \ldots, p_g)$-Würfeln

Oder: Was $2$ recht ist, soll $g$ billig sein!

(vgl. Buch S. 28)

Definition ("$n$-faches $(p_1, \ldots, p_g)$-Würfeln"):

Seien $g \in \mathbb{N}$ und $p_1, \ldots, p_g \geq 0$ mit $p_1 + \ldots + p_g = 1$.

Wir definieren Verteilungsgewichte auf

$$S := \{a = (a_1, \ldots, a_n) \ : \ a_i \in \{1, \ldots, g\}\}$$

durch

$$\rho(a_1, \ldots, a_n) := p_{a_1} \cdot p_{a_2} \cdots p_{a_n}.$$

Eine Zufallsvariable $W$ mit diesem Zielbereich $S$

und diesen Verteilungsgewichten $\rho$ nennen wir

$n$-faches $(p_1, \ldots, p_g)$-Würfeln.

Für jedes $a \in S$ mit

$k_1$ Komponenten gleich 1,

$k_2$ Komponenten gleich 2,

...

$k_g$ Komponenten gleich $g$

ist dann

$$\mathbf{P}(W = a) = p_1^{k_1} p_2^{k_2} \cdots p_g^{k_g}$$

# 9. Vom $(p_1, \ldots, p_g)$-Würfeln zur Multinomialverteilung

$W = (W_1, \ldots, W_n)$ sei ein $n$-faches $(p_1, \ldots, p_g)$-Würfeln

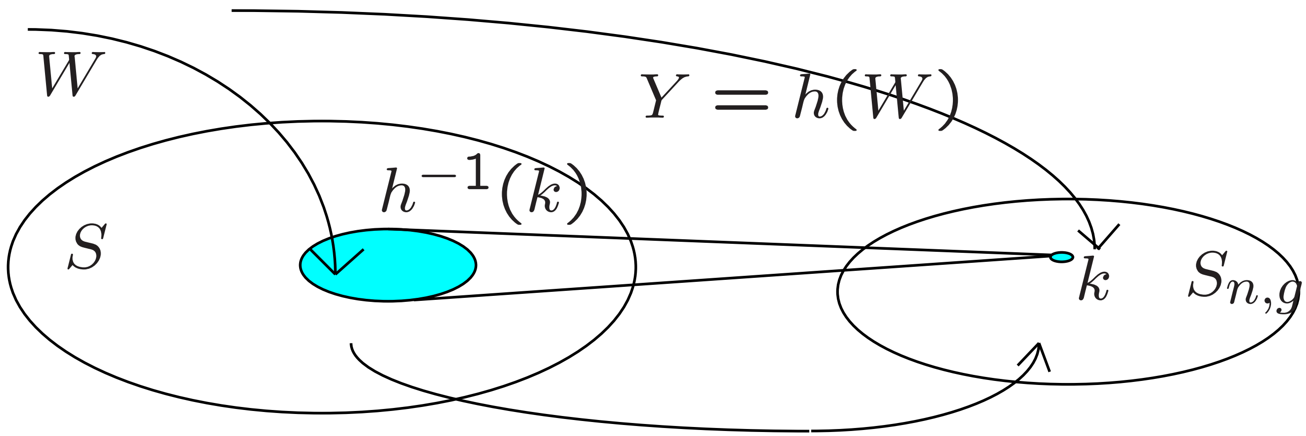
$$Y_j := \#\{i \ : \ W_i = j\}$$

(die Anzahl der Würfe mit Ergebnis $j$).

$Y := (Y_1, \ldots, Y_g)$ hat dann den Zielbereich

$$S_{n,g} = \{(k_1, \ldots, k_g) \ : k_1 + \ldots + k_g = n\}.$$

Verteilung von $Y$ = ?




$$h(a_1, \ldots, a_n) = (k_1, \ldots, k_g) =: k$$

mit $k_j := \#\{i : a_i = j\}),\ j = 1, \ldots, g$

Jedes $a \in S$ mit $h(a) = (k_1, \ldots, k_g)$

hat Gewicht $p_1^{k_1} \ldots p_g^{k_g}$

Wieviele solche $a$ gibt es?

Dazu überlegen wir:

Auf wieviele Arten kann man

$n$ Objekte so auf $g$ Fächer verteilen,

dass das $j$-te Fach genau $k_j$ Objekte enthält?

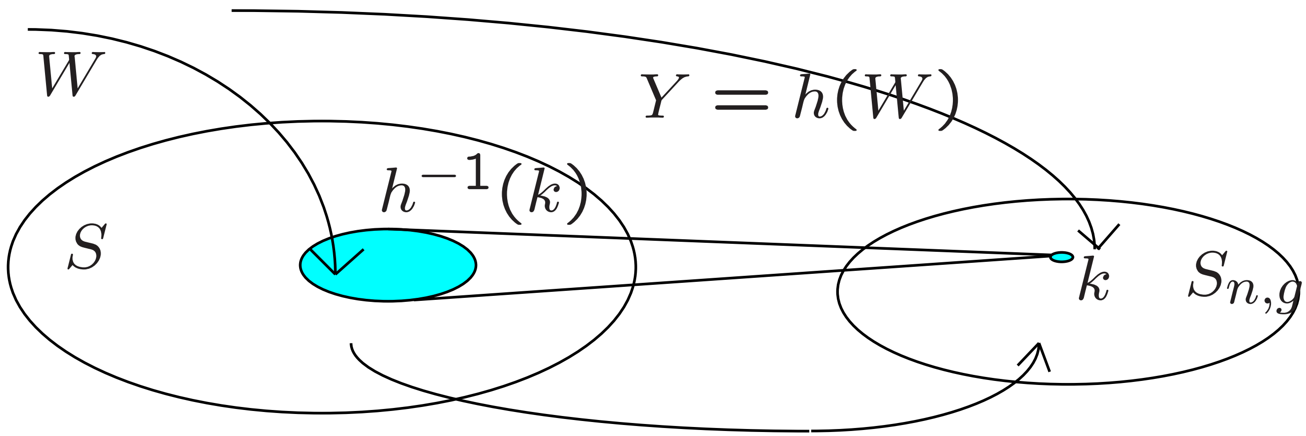
Dabei ist $k_1 + \cdots + k_g = n$.

Die Antwort ist:

$$\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdots \binom{n-k_1-\cdots-k_{g-1}}{k_g}$$

$$= \frac{n!}{k_1!\,k_2!\cdots k_g!} =: \binom{n}{k_1,\ldots,k_g}$$

Multinomialkoeffizient, lies: $n$ über $k_1, \ldots, k_g$

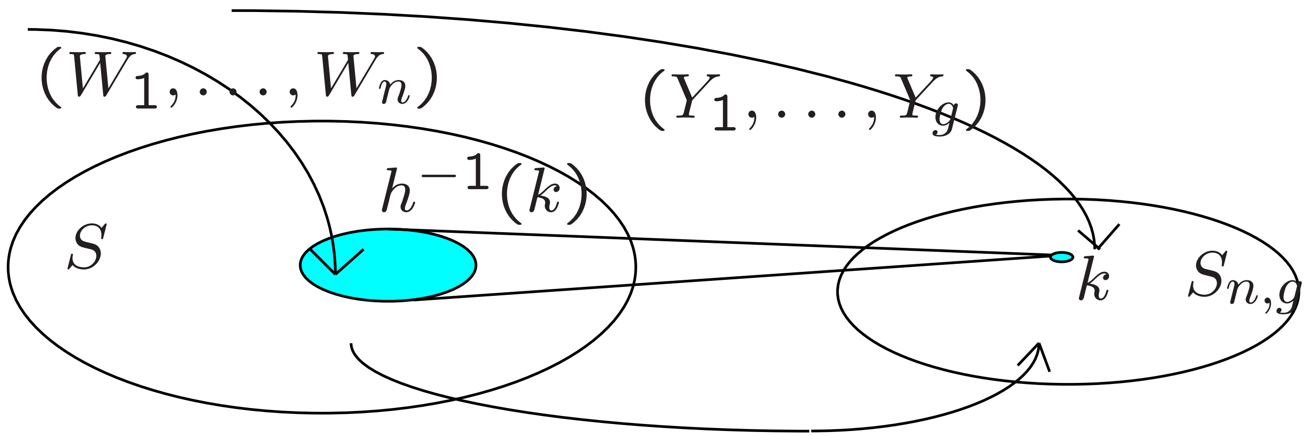



$$h(a_1, \ldots, a_n) = (k_1, \ldots, k_g) = k$$

Jedes $a \in S$ mit $h(a) = (k_1, \ldots, k_g)$

hat Gewicht $p_1^{k_1} \ldots p_g^{k_g}$

Wieviele solche $a$ gibt es?



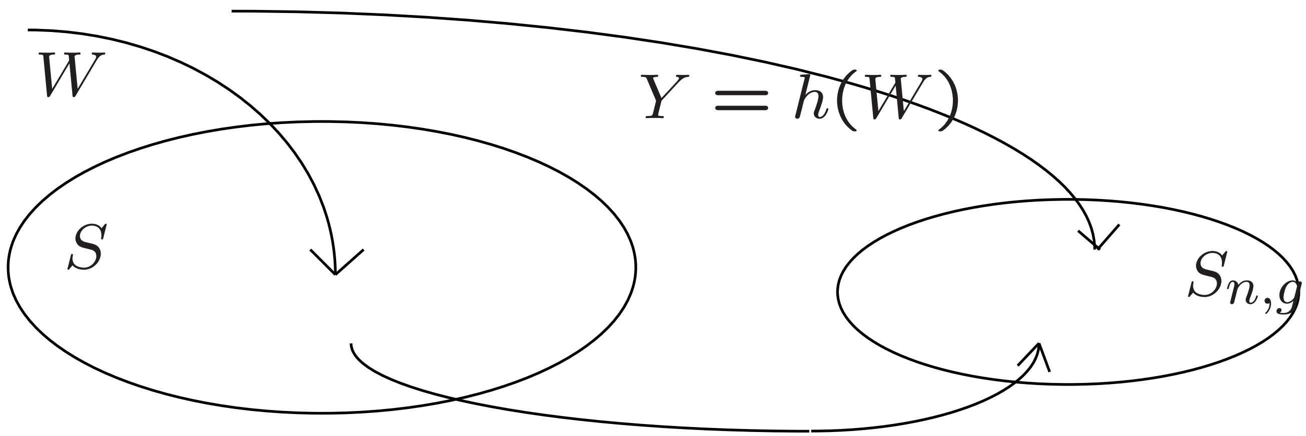

$$h(a_1, \ldots, a_n) = (k_1, \ldots, k_g)$$

Jedes $a \in S$ mit $h(a) = (k_1, \ldots, k_g)$

hat Gewicht $p_1^{k_1} \ldots p_g^{k_g}$

Es gibt $\binom{n}{k_1, \ldots, k_g}$ solche $a$.

$$\mathbf{P}(Y_1 = k_1, \ldots, Y_g = k_g) = \binom{n}{k_1, \ldots, k_g} p_1^{k_1} \ldots p_g^{k_g}$$




$$h(a_1, \ldots, a_n) = (k_1, \ldots, k_g) =: k$$

mit $k_j := \#\{i : a_i = j\}, \ j = 1, \ldots, g$

$$\mathbf{P}(Y_1 = k_1, \ldots, Y_g = k_g) = \binom{n}{k_1, \ldots, k_g} p_1^{k_1} \ldots p_g^{k_g}$$

Definition:

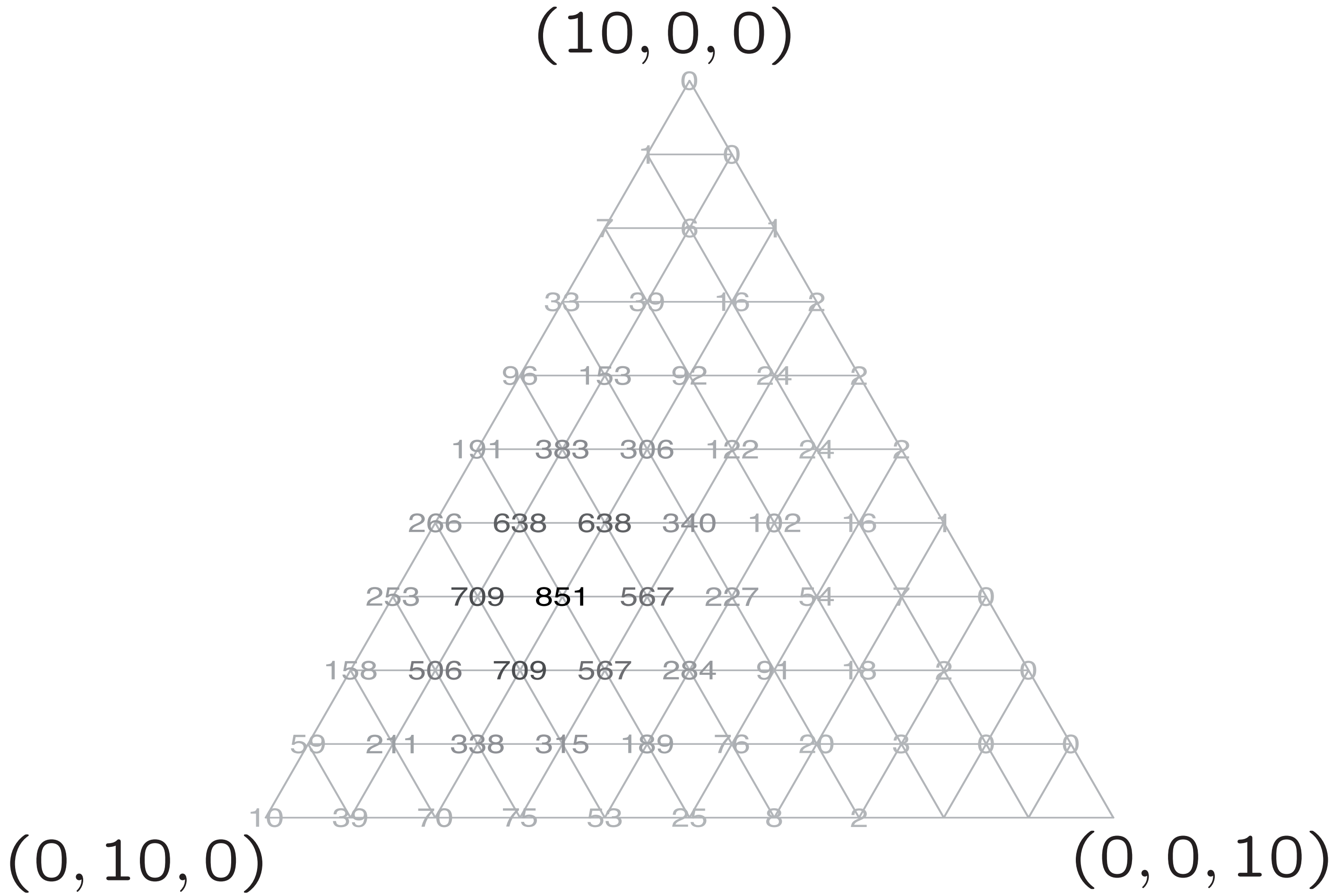
Eine Zufallsvariable $X$ mit Zielbereich $S_{n,g}$

heißt *multinomialverteilt* mit Parametern $n; p_1, \ldots, p_g$,

wenn

$$\mathbf{P}(X = (k_1, \ldots, k_g)) = \binom{n}{k_1, \ldots, k_g} p_1^{k_1} \ldots p_g^{k_g},$$

$$(k_1, \ldots, k_g) \in S_{n,g} .$$




Gewichte der Multinomialverteilung, notiert in Vielfachen von $\frac{1}{10000}$,
für $n = 10$, $g = 3$, $p_1 = 0.3$, $p_2 = 0.5$, $p_3 = 0.2$

## Resumé

| $g \in \mathbb{N}$ | $g = 2$ |
|---|---|
| Würfeln | Münzwurf |
| Besetzung der Ausgänge | Anzahl der Erfolge |
| Multinomialverteilung | Binomialverteilung |

Zwei wichtige Beispiele aus der heutigen Vorlesung

Stichprobenziehen $\to$ Münzwurf $\to$ Binomialverteilung

$(g = 2)$

Stichprobenziehen $\to$ “Würfeln” $\to$ Multinomialverteilung

$(g = 3)$

werden illustriert durch die

Wandtner’schen R-Programme zu VL 2b.