

Kapitel 4

Diskrete Verteilungen und Zufallsgrößen

Wahrscheinlichkeitsverteilungen und Zufallsgrößen haben wir in dem sehr allgemeinen Rahmen von Wahrscheinlichkeitsräumen $(\Omega, \mathfrak{A}, P)$ eingeführt. In dieser Allgemeinheit, die den Vorteil der begrifflichen Klarheit, Übersichtlichkeit und der Spezialisierungsmöglichkeit hat, ist jedoch eine detaillierte Untersuchung bzw. Ausgestaltung der mit ihnen zusammenhängenden Begriffe anspruchsvoll und bedarf der Kenntnis der Maßtheorie. Für viele Anwendungen ist diese Allgemeinheit aber nicht notwendig. Wir stellen sie also zunächst zurück und schränken uns in diesem Kapitel auf den Spezialfall diskreter Wahrscheinlichkeitsverteilungen ein.

In diesem Fall tritt die Maßtheorie in den Hintergrund, da man es im Grunde stets mit höchstens abzählbar unendlich vielen Versuchsausgängen bzw. möglichen Werten (bei Zufallsgrößen) zu tun hat und deshalb der Verwendung der Potenzmenge als relevante σ -Algebra von Teilmengen nichts im Wege steht.

Diskrete Verteilungen sind, grob gesprochen, solche, bei denen die "Wahrscheinlichkeitsmasse" in höchstens abzählbar vielen Punkten konzentriert ist.

4.1 Definitionen und Beispiele

Es seien $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und I eine Teilmenge der Menge N_0 aller natürlichen Zahlen.

Definition 4.1 Die Wahrscheinlichkeitsverteilung P heißt eine diskrete Ver-

teilung, falls es eine höchstens abzählbare Menge $\Omega_P := \{\omega_i : i \in I\}$ aus Ω gibt mit $\{\omega_i\} \in \mathfrak{A}$, $i \in I$, und $P(\Omega \setminus \Omega_P) = 0$.

Insbesondere ist jede Wahrscheinlichkeitsverteilung P auf (Ω, \mathfrak{A}) diskret, falls Ω selbst höchstens abzählbar unendlich ist.

Folgerungen 4.2 Mit der Bezeichnung $p_i := P(\{\omega_i\})$, $i \in I$, gilt

1.

$$\sum_{i \in I} p_i = 1 \quad (4.1)$$

2. Für alle A aus \mathfrak{A} ist

$$P(A) = P(A \cap \Omega_P) = P(\{\omega_i \in \Omega_P : \omega_i \in A\}) = \sum_{i: \omega_i \in A} p_i. \quad (4.2)$$

Das bedeutet, jede diskrete Wahrscheinlichkeitsverteilung P ist durch Angabe der Paare $(\omega_i, p_i)_{i \in I}$ eindeutig bestimmt.

Aus diesem Grund wird häufig die Folge $((\omega_i, p_i), i \in I)$ bereits als diskrete Wahrscheinlichkeitsverteilung auf $\Omega_P = \{\omega_i : i \in I\}$ bezeichnet. Die Zahlen p_i heißen *Einzelwahrscheinlichkeiten* der Verteilung P .

3. o.B.d.A. kann man $p_i > 0, i \in I$, annehmen. Gilt nämlich $p_i = 0$ für ein $i \in I$, so entfernt man dieses ω_i aus Ω_P . Die Menge $\Omega_P^{\min} := \{\omega_i \mid i \in I, p_i > 0\}$ heißt *Träger* der diskreten Verteilung P .

Die Formel (4.2) kann man nutzen, um P für jede Teilmenge A von Ω zu definieren, nicht nur für $A \in \mathfrak{A}$. Bei diskreten Verteilungen P ist also immer eine Erweiterung von \mathfrak{A} auf $\mathfrak{P}(\Omega)$ möglich. Wir setzen in Zukunft deshalb bei diskreten Verteilungen stets voraus, dass $\mathfrak{A} = \mathfrak{P}(\Omega)$ gilt.

Beispiel 4.3

a) Gibt es Elemente $\omega_1, \dots, \omega_N$ mit $P(\{\omega_k\}) = p_k = \frac{1}{N}$, so spricht man von der "Gleichmäßigen diskrete Verteilung auf $\{\omega_1, \dots, \omega_N\}$."

b) Gibt es ein $\omega_0 \in \Omega$ mit $P(\{\omega_0\}) = 1$, so heißt P die "ausgeartete Verteilung, konzentriert in ω_0 " oder die in ω_0 konzentrierte Einpunktverteilung.

c) Die Binomialverteilung

Es seien $n \in N_1 = \{1, 2, \dots, m, \dots\}$ und $p \in (0, 1)$. Durch

$$b(n, p; k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}$$

sind die Einzelwahrscheinlichkeiten einer Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$ gegeben. Diese Verteilung heißt Binomialverteilung mit den Parametern n und p .

d) Die Poissonverteilung

Es sei $\lambda > 0$. Durch

$$p_k(\lambda) := \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0$$

sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf $N_0 = \{0, 1, 2, \dots, k, \dots\}$ gegeben.

Diese heißt Poissonverteilung mit dem Parameter λ .

e) Die geometrische Verteilung

Es sei $p \in (0, 1)$. Durch

$$g_k(p) := (1-p)^k p, \quad k \geq 0$$

sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf $N_0 \setminus \{0\} = \{1, 2, \dots, k, \dots\}$ gegeben. Diese Verteilung heißt geometrische Verteilung mit dem Parameter p .

f) Die hypergeometrische Verteilung

Es seien R, S positive ganze Zahlen, $M := R + S$ und m eine ganze Zahl mit $1 \leq m \leq M$. Durch

$$h(M, R, m; k) := \frac{\binom{R}{k} \binom{S}{m-k}}{\binom{M}{m}}$$

sind die Einzelwahrscheinlichkeiten einer Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, M\}$ gegeben. Diese Verteilung heißt hypergeometrische Verteilung mit den Parametern M, R, m .

Es gilt $h(M, R, m; k) > 0$ genau dann, wenn $[\max(0, m - S) \leq k \leq \min(m, R)]$, wie man leicht an der Definition der Binomialkoeffizienten erkennt.

- g) Die negative Binomialverteilung
Es seien $p \in (0, 1)$ und $v > 0$. Durch

$$NB(p, v; k) := \binom{-v}{k} (-q)^k p^v \quad , \quad k \geq 0$$

mit $q = 1 - p$ sind die Einzelwahrscheinlichkeiten einer diskreten Verteilung auf $\{0, 1, 2, \dots, k, \dots\}$ gegeben. Diese Verteilung heißt negative Binomialverteilung mit den Parametern p und v .

Man beachte:

$$\binom{-v}{k} := \frac{(-v)(-v-1)\dots(-v-k+1)}{k!} = (-1)^k \binom{v+k-1}{k}$$

Die hier vorgestellten diskreten Verteilungen treten in Theorie und Anwendungen der Stochastik häufig auf. Sie sind Bestandteil gewisser Standardmodelle der Wahrscheinlichkeitstheorie und teilweise durch Grenzübergänge miteinander verbunden. Exemplarisch konstruieren wir als erstes ein Modell, bei dem die hypergeometrische Verteilung vorkommt und geben dann zwei Grenzwertaussagen an, die die hypergeometrische, die Binomial- und die Poissonverteilung miteinander verbinden. Zunächst erweitern wir jedoch den Begriff der diskreten Verteilung auf Zufallsgrößen.

Definition 4.4 Ist X eine Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit Werten in (E, \mathfrak{E}) , so heißt X eine diskret verteilte Zufallsgröße, kurz: diskrete Zufallsgröße, falls ihre Wahrscheinlichkeitsverteilung P^X auf (E, \mathfrak{E}) diskret ist.

In diesem Fall gibt es nach Definition eine Folge $(x_i, i \in I)$ mit $I \subseteq N_0$ von Elementen aus E mit

$$\sum_{i \in I} P^X(\{x_i\}) = \sum_{i \in I} P(X = x_i) = 1 \text{ und} \quad (4.3)$$

$$P^X(B) = \sum_{i \in I: x_i \in B} P(X = x_i), \quad B \in \mathfrak{E}. \quad (4.4)$$

Verteilungsfunktionen diskreter Verteilungen auf R_1

Es seien $(x_i, i \in I)$ eine Folge reeller Zahlen und $((x_i, p_i), i \in I)$ eine diskrete Wahrscheinlichkeitsverteilung. Das von ihr erzeugte Wahrscheinlichkeitsmaß P hat die Form

$$P(A) = \sum_{i: x_i \in A} p_i, \quad A \subseteq R_1$$

(siehe Formel (4.2)).

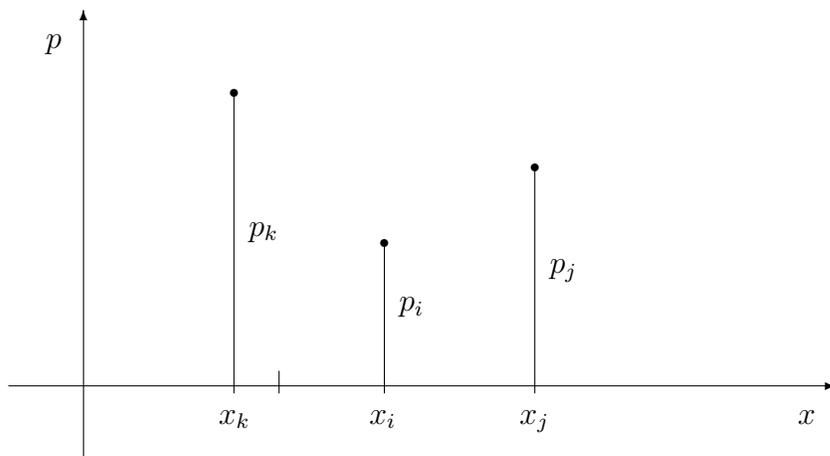


Bild 4.1

Die *Verteilungsfunktion* F der diskreten Verteilung $((x_i, p_i), i \in I)$ ist definiert durch (siehe (3.27))

$$F(x) := P((-\infty, x]) = \sum_{i: x_i \leq x} p_i, \quad x \in R_1. \quad (4.5)$$

Für die Funktion F gilt die Aussage 3.31. Außerdem haben wir die

Aussage 4.5 *Die Verteilungsfunktion F hat folgende Eigenschaften:*

- ΔF ist konstant auf jedem Intervall $[a, b)$, das keine der Zahlen x_i im Inneren enthält.
- $F(x_i) - F(x_i - 0) = p_i$, $i \in I$

Der Beweis folgt unmittelbar aus der Definition (4.3).

Funktionen diskret verteilter Zufallsgrößen

Es sei X eine diskret verteilte Zufallsgröße mit der Menge der möglichen Werte $E = \{x_i : i \in I\}$ und den zugehörigen Einzelwahrscheinlichkeiten $(p_i^X, i \in I)$. Ist ψ eine Funktion von E in eine abzählbare Menge $F = \{f_j : j \in \mathcal{J}\}$, so ist die Zufallsgröße $Y := \psi(X)$ ebenfalls diskret verteilt.

Aussage 4.6 *Die Verteilung der Zufallsgröße $Y = \psi(X)$ ist diskret. Ihre möglichen Werte sind die Elemente von $F = \{\psi(x_i) : i \in I\} = \{f_j : j \in \mathcal{J}\}$ mit den Einzelwahrscheinlichkeiten*

$$p_j^Y = \sum_{\substack{i \in I: \\ \psi(x_i) = f_j}} p_i^X, \quad j \in \mathcal{J} \quad (4.6)$$

Beweis: $p_j^Y = P^Y(\{f_j\}) = P^X(\psi^{-1}(\{f_j\})) = \sum_{\substack{i \in I: \\ \psi(x_i) = f_j}} p_i^X. \quad \square$

4.2 Die hypergeometrische Verteilung

Das folgende Modell steht für viele Situationen, in denen eine zufällige Auswahl von Elementen aus einer aus zwei Typen von Elementen bestehenden Menge (ohne Zurücklegen) vorgenommen wird (Lotto "6 aus 49", Qualitätskontrolle mit Hilfe einer Stichprobe usw.).

Gegeben sei eine Urne mit M Kugeln, davon R rote und S schwarze:

$$M = R + S.$$

Die Kugeln seien durchnummeriert von 1 bis M , dabei mögen die roten Kugeln die Nummern 1 bis R tragen. Auf gut Glück werden m Kugeln ausgewählt, nacheinander, ohne Zurücklegen. Der Einfachheit halber setzen wir $m \leq \min(R, S)$ voraus.

Die möglichen Ausgänge ω dieses Versuches sind, wenn die Reihenfolge der ausgewählten Kugeln keine Rolle spielt, m -elementige Teilmengen von $\{1, 2, \dots, M\}$:

$$\omega = \{i_1, \dots, i_m\}, \quad i_k \in \{1, 2, \dots, M\}, \quad k = 1, \dots, m.$$

Die Menge Ω aller dieser ω hat $\binom{M}{m}$ Elemente. Es gibt also $N = \binom{M}{m}$ mögliche Versuchsausgänge.

Weil die Auswahl auf gut Glück erfolgte, hat jedes $\omega \in \Omega$ die gleiche Wahrscheinlichkeit aufzutreten. Folglich haben wir ein Laplace-Experiment mit dem Parameter N :

$$P(\{\omega\}) = \frac{1}{N} \binom{M}{m}^{-1}, \quad \omega \in \Omega.$$

Die Zufallsgröße X , definiert durch

$$X(\omega) = \sum_{k=1}^m \mathbb{1}_{\{1, \dots, R\}}(i_k), \quad \omega = \{i_1, i_2, \dots, i_m\} \in \Omega,$$

gibt an, wieviel rote Kugeln in der "Stichprobe" ω enthalten sind. Sie hat die möglichen Werte $0, 1, \dots, m$, und für ihre Einzelwahrscheinlichkeiten gilt

$$P(X = j) = \frac{\#\{\omega : X(\omega) = j\}}{N} = \frac{\binom{R}{j} \binom{M-R}{m-j}}{\binom{M}{m}}, \quad j = 0, 1, \dots, m. \quad (4.7)$$

Es gilt somit

Aussage 4.7 Werden aus einer Urne mit R roten und S schwarzen Kugeln m Kugeln nacheinander, ohne Zurücklegen und auf gut Glück ausgewählt, so hat die Zufallsgröße X , die die Anzahl der roten Kugeln in der ausgewählten Stichprobe angibt, eine hypergeometrische Verteilung mit den Parametern $M = R + S$, R und m . Es gilt also (4.7).

Bemerkung 4.8 Die Formel (4.7) bleibt auch gültig, falls $m > \min(R, M - R)$ gilt.

Beispiel 4.9 (Lotto "6 aus 49")

$$M = 49, m = 6, R = 6$$

(rote Kugeln $\hat{=}$ Zahlen auf dem Tippschein,

schwarze Kugeln $\hat{=}$ restliche der 49 Zahlen)

X = Zahl der auf dem Tippschein richtig getippten Zahlen.

$$P(X = k) = \frac{\binom{6}{k} \binom{43}{6-k}}{\binom{49}{6}}, k = 0, 1, \dots, 6.$$

k	0	1	2	3
$P(X = k)$	0,43596498	0,41301945	0,13237803	0,0176504

k	4	5	6
$P(X = k)$	0,00096862	$1,845 \cdot 10^{-5}$	$7,15 \cdot 10^{-8}$

Aussage 4.10 Mit der Bezeichnung

$$h(M, R, m; k) = \frac{\binom{R}{k}}{\binom{M-R}{m-k}}, k = 0, \dots, m, \quad (4.8)$$

gilt

$$\lim_{\substack{M, r \rightarrow \infty \\ R, M \rightarrow \infty}} h(M, R, m; k) = \binom{m}{k} p^k (1-p)^{m-k}, \quad (4.9)$$

wobei der Limes derart gebildet wird, dass für gegebenes p aus $(0, 1)$ gilt $M \rightarrow \infty$, $R \rightarrow \infty$ mit $R/M \rightarrow p$, m und k bleiben fest.

Im Grenzfall geht die hypergeometrische Verteilung also unter den genannten Bedingungen in eine Binomialverteilung mit den Parametern (m, p) über.

Beweis: Als Übungsaufgabe. (Man beachte, dass m und k beim Grenzübergang festgehalten werden.)

Satz 4.11 (Poissonscher Grenzwertsatz)

Es gilt für jedes $\lambda > 0$

$$\lim_{\substack{m \rightarrow \infty \\ m \cdot p_m \rightarrow \lambda}} \binom{m}{k} p_m^k (1 - p_m)^{m-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0 \quad (4.10)$$

Beweis: Wir schreiben $\binom{m}{k} p_m^k (1 - p_m)^{m-k}$ in der Form $\frac{1}{k!} \left(\prod_{j=0}^{k-1} (m - j) p_m \right) \cdot \left(1 - \frac{p_m \cdot m}{m} \right)^m \cdot \frac{1}{(1 - p_m)^k}$. Wegen $\prod_{j=0}^{k-1} (m - j) p_m \rightarrow \lambda^k$, $\left(1 - \frac{p_m \cdot m}{m} \right)^m \rightarrow e^{-\lambda}$ und $(1 - p_m)^k \rightarrow 1$ für $m \rightarrow \infty$ mit $m p_m \rightarrow \lambda$ folgt die Behauptung. \square

4.3 Erwartungswert und Varianz

Erwartungswert und Varianz sind aufschlussreiche Kenngrößen einer Wahrscheinlichkeitsverteilung. Sie geben Anhaltspunkte dafür, um welchen "Schwerpunkt" sich die Trägerpunkte der Verteilung gruppieren bzw. wie stark sie um diesem Schwerpunkt "streuen".

Erwartungswert

Es sei $((x_i, p_i), i \in I \subseteq N_0)$ eine diskrete Wahrscheinlichkeitsverteilung auf R_1 . Ein zufälliger Versuch werde n mal (jedes Mal neu, unter im Wesentlichen gleichen Bedingungen) ausgeführt und zwar so, dass der Wert x_i mit der

Wahrscheinlichkeit p_i erscheint. Als Ergebnis der Versuchsreihe erhalten wir eine Folge (y_1, \dots, y_n) von Versuchsausgängen, wobei jedes $y_j, j = 1, 2, \dots, n$, gleich einem der $x_i, i \in I$, ist. Es sei n_i die (absolute) Häufigkeit, mit der x_i als Versuchsausgang unter den y_1, \dots, y_n auftritt, in einer Formel ausgedrückt, heißt das

$$n_i = \sum_{k=1}^n \mathbb{1}_{\{x_i\}}(y_k).$$

Offenbar gilt $\sum_{i \in I} n_i = n$ und $\sum_{i \in I} n_i x_i = \sum_{j=1}^n y_j$.

Angenommen, wir erhalten nach jeder Versuchsdurchführung von einem Veranstalter so viele Euro, wie der Versuchsausgang x_i als Zahl angibt (negative Werte bedeuten Zahlungsverpflichtung für uns), dann haben wir insgesamt

$$\sum_{j=1}^n y_j = \sum_{i \in I} n_i x_i \text{ Euro bekommen. Pro Versuch sind das also im Durchschnitt}$$

$\frac{1}{n} \sum_{k=1}^n y_j = \sum_{i \in I} \frac{n_i}{n} x_i$. Wir erinnern uns, bei großer Versuchsanzahl n ist die relative Häufigkeit $\frac{n_i}{n}$ etwa gleich der Wahrscheinlichkeit p_i (Empirisches Gesetz der großen Zahlen).

Der Wert $\mu := \sum_{i \in I} p_i x_i$ gibt also näherungsweise den Geldbetrag in Euro an, den wir in einer langen Reihe von Versuchen pro Versuch erhalten, wir sagen, den wir pro Versuch zu *erwarten* haben.

Dieser Wert wäre auch der *faire Preis*, den wir vor Durchführung jedes Versuches an den Veranstalter zu bezahlen hätten.

Definition 4.12 Der Erwartungswert μ einer diskreten Verteilung $((x_i, p_i), i \in I)$ mit $x_i \in \mathbb{R}_1, i \in I$, existiert und ist definiert als

$$\mu = \sum_{i \in I} x_i p_i, \quad \text{falls} \quad \sum_{i \in I} x_i^+ p_i < \infty \quad \text{oder} \quad \sum_{i \in I} x_i^- p_i < \infty.$$

Anderenfalls sagt man, $((x_i, p_i), i \in I)$ besitze keinen Erwartungswert.

Gilt $\sum_{i \in I} |x_i| p_i < \infty$, so ist $|\mu| < \infty$. In diesem Fall sagt man, die Verteilung hat einen endlichen Erwartungswert. (Dabei ist $x^+ = \max(x, 0)$, $x^- = \max(-x, 0)$. Es gilt $x = x^+ - x^-$, $|x| = x^+ + x^-$.)

Das empirische Gesetz der großen Zahlen kann man nach diesen Überlegungen also auch für arithmetische Mittel formulieren:

Wenn der Erwartungswert μ existiert, so nähert sich das arithmetische Mittel $\frac{1}{n} \sum_{k=1}^n y_j$ der Versuchsergebnisse immer mehr diesem Erwartungswert.

Fasst man die Verteilung $((x_i, p_i), i \in I)$, als eine Verteilung von Gewichten der Masse p_i im Punkt $x_i, i \in I$, auf, so ist der Erwartungswert μ der physikalische Schwerpunkt dieser Massenverteilung. Um ihn gruppieren sich die möglichen Werte x_i der Verteilung. In erster Näherung liefert also μ Informationen über die "Lage" dieser Verteilung. Man bezeichnet deshalb μ auch als *Lageparameter*. Eine Verteilung heißt *zentriert*, falls ihr Erwartungswert μ gleich Null ist.

Verschiebt man jeden Punkt x_i um einen Wert a in $x_i + a$, so verschiebt sich auch der Erwartungswert μ um a in den neuen Erwartungswert $\mu + a$.

Setzt man $a = -\mu$, ergibt sich als neue Verteilung $((x_i - \mu, p_i), i \in I)$, und deren Erwartungswert ist gleich Null. Sie ist also *zentriert*.

Beispiel 4.13 (Erste Fortsetzung des Beispiels 4.3):

$$\text{a) } \mu = \frac{1}{N} \sum_{k=1}^N \omega_k, \quad \text{falls } \Omega \subseteq R_1$$

$$\text{b) } \mu = \omega_0, \quad \text{falls } \Omega \subseteq R_1$$

$$\text{c) } \mu = np$$

$$\text{d) } \mu = \lambda$$

$$\text{e) } \mu = \frac{1-p}{p}$$

$$\text{f) } \mu = \frac{Rm}{M}$$

$$\text{g) } \mu = v \cdot \frac{1-p}{p}$$

Definition 4.14 Ist X eine diskret verteilte reellwertige Zufallsgröße, so bezeichnet man als Erwartungswert von X den Erwartungswert ihrer Wahrscheinlichkeitsverteilung P^X und verwendet für ihn das Symbol EX :

$$EX = \sum_{i \in I} x_i P^X(\{x_i\}) = \sum_{i \in I} x_i P(X = x_i)$$

Dabei bilden die $x_i, i \in I$, die möglichen Werte von X .

Eine sehr einfache Zufallsgröße ist $X(\omega) = \mathbb{1}_A(\omega)$ mit $A \in \mathfrak{A}$. Es gilt $EX = E\mathbb{1}_A = P(A)$.

Aussage 4.15 (Erwartungswert der Funktion einer Zufallsgröße)

Es sei X eine diskret verteilte Zufallsgröße über einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Werten in irgendeiner abzählbaren Menge $E = \{x_i : i \in I \subseteq \mathbb{N}_0\} \subseteq \mathbb{R}_1$ und mit den entsprechenden Einzelwahrscheinlichkeiten $(p_i^X, i \in I)$. Außerdem sei ψ eine reellwertige Funktion auf E mit Werten in $F = \{f_j : j \in \mathfrak{J} \subseteq \mathbb{N}_0\}$. Dann ist $Y = \psi(X)$ eine reellwertige diskret verteilte Zufallsgröße, und es gilt (siehe (5)):

$$EY = E\psi(X)$$

$$\sum_{i \in I} \psi(x_i) P(X = x_i) p_i^X \tag{4.11}$$

wobei dieser Erwartungswert nach Definition nicht existiert, falls

$$\sum_{i \in I} (\psi(x_i))^+ P(X = x_i) \text{ und } \sum_{i \in I} (\psi(x_i))^- P(X = x_i) = \infty \text{ gilt.}$$

Beweis:

$$EY = \sum f_j p_j^Y = \sum_j f_j \sum_{\substack{i \in \mathcal{J}: \\ \psi(x_i) = f_j}} p_i^X = \sum_{j \in \mathcal{J}} \sum_{\substack{i \in \mathcal{J}: \\ \psi(x_i) = f_j}} f_j p_i^X = \sum_{i \in \mathcal{J}} \psi(x_i) p_i^X.$$

Beispiel 4.16

- 1) Ist $\psi(x) = ax + b$, $x \in R_1$, a, b reellwertige Konstanten, so gilt, sofern EX existiert,

$$E(aX + b) = a(EX) + b$$

- 2) Für jede reellwertige diskrete Zufallsgröße X ist auch X^2 eine Zufallsgröße, und es gilt

$$EX^2 = \sum_{i \in I} x_i^2 P(X = x_i).$$

Momente diskreter Verteilungen auf R_1

Es sei $((x_i, p_i), i \in I)$, eine diskrete Wahrscheinlichkeitsverteilung auf R_1 .

Definition 4.17 *Es sei $k \geq 1$. Als k -tes Moment der Wahrscheinlichkeitsverteilung $((x_i, p_i), i \in I)$, bezeichnet man die Größe*

$$\mu_k := \sum_{i \in I} x_i^k p_i,$$

sofern $\sum (x_i^+)^k p_i < \infty$ oder $\sum (x_i^-)^k p_i < \infty$. Anderenfalls sagt man, falls k ungerade ist, das k -te Moment existiert nicht. Sind beide Summen endlich, so konvergiert die Summe $\sum_{i \in I} |x_i|^k p_i$ und das k -te Moment $\mu_k = \sum_{i \in I} x_i^k p_i$ ist endlich.

Der Erwartungswert ist offensichtlich das erste Moment der Verteilung (x_i, p_i) : $\mu = \mu_1$. Gilt $|\mu_k| < \infty$ für ein $k > 1$, so ist auch $|\mu_l| < \infty$ für alle l mit $1 \leq l < k$. Das folgt sofort aus $|\mu_l| \leq \sum_{i \in I} |x_i|^l p_i \leq \sum [\max(1, |x_i|)]^k p_i \leq 1 + \sum_{i \in I} |x_i|^k p_i$.

Definition 4.18 *Es sei $k \geq 2$. Als k -tes zentrales Moment einer Wahrscheinlichkeitsverteilung (x_i, p_i) , $i \in I$, bezeichnet man das k -te Moment der zentrierten Verteilung $(x_i - \mu, p_i)$, $i \in I$:*

$$m_k := \sum_{i \in I} (x_i - \mu)^k p_i,$$

sofern $\sum ((x_i - \mu)^+)^k p_i < \infty$ oder $\sum ((x_i - \mu)^-)^k p_i < \infty$ gilt. Anderenfalls sagt man, falls k ungerade ist, das k -te zentrale Moment existiert nicht.

Es gilt: $|m_k| < \infty$ genau dann, wenn $|\mu_k| < \infty$ ($k \geq 2$). In diesem Fall ist

$$m_k = \sum_{\ell=0}^k \binom{k}{\ell} \mu_\ell (-\mu)^{k-\ell}, \quad k \geq 2 \quad (4.12)$$

mit $\mu_0 := 1$, insbesondere gilt:

$$m_2 = \mu_2 - \mu_1^2. \quad (4.13)$$

Umgekehrt haben wir

$$\mu_k := \sum_{i \in I} (x_i - \mu + \mu)^k p_i = \sum_{\ell=0}^k \binom{k}{\ell} m_\ell \cdot \mu^{k-\ell} \quad (4.14)$$

mit $m_0 := 1, m_1 = 0$.

Mit Hilfe der Momente einer Wahrscheinlichkeitsverteilung auf R_1 kann man eine erste Vorstellung von der Lage und der Form der Wahrscheinlichkeitsverteilung auf R_1 gewinnen.

Definition 4.19 Als k -tes Moment einer diskreten reellwertigen Zufallsgröße X über $(\Omega, \mathfrak{A}, P)$ bezeichnet man das k -te Moment μ_k^X ihrer Wahrscheinlichkeitsverteilung P^X .

Es gilt:

$$\mu_k^X = \sum_{i \in I} x_i^k P^X(\{x_i\}) = \sum_{i \in I} x_i^k P(X = x_i) = E(X^k). \quad (4.15)$$

mit den gleichen Existenz- bzw. Nichtexistenzbedingungen wie beim k -ten Moment irgendeiner diskreten Verteilung auf R_1 . Wir schreiben $\mu^X = \mu_1^X$. Schließlich führt man für $k \geq 2$ das k -te zentrale Moment für X ein als

$$\begin{aligned} m_k^X &= \sum_{i \in I} (x_i - \mu^X)^k P^X(\{x_i\}) = \\ &= \sum_{i \in I} (x_i - \mu^X)^k P(X = x_i) = E(X - \mu^X)^k. \end{aligned} \quad (4.16)$$

Varianz

Das erste Moment, der Erwartungswert μ , kennzeichnet die Lage der Verteilung, das zweite zentrale Moment vermittelt eine Vorstellung, wie breit die Verteilung um den Erwartungswert platziert ist. Es hat einen eigenen Namen.

Definition 4.20 Als Varianz oder Streuung der Wahrscheinlichkeitsverteilung $((x_i, p_i), i \in I)$ bezeichnet man die Größe

$$\sigma^2 := \sum_{i \in I} (x_i - \mu)^2 p_i. \quad (4.17)$$

Die Wurzel aus der Varianz $\sigma = (\sigma^2)^{\frac{1}{2}}$ nennt man Standardabweichung der zugrunde liegenden Verteilung.

Es gilt $\sigma^2 \geq 0$. Wir haben $\sigma^2 = 0$ genau dann, wenn die Verteilung $((x_i, p_i), i \in I)$ ausgeartet ist, also die Verteilung in nur einem Punkt x_{i_0} für ein $i_0 \in I$ konzentriert ist, d. h. wenn gilt $p_{i_0} = 1$. In diesem Fall ist $\mu = x_{i_0}$.

Beispiel 4.21 (Zweite Fortsetzung der Beispiele aus 4.3):

$$\text{a) } \sigma^2 = \frac{1}{N} \sum_{i \in I} (\omega_i - \mu)^2, \text{ falls } \Omega \subseteq R_1$$

$$\text{b) } \sigma^2 = 0 \quad , \text{ falls } \omega_0 \in R_1$$

$$\text{c) } \sigma^2 = np(1-p)$$

$$\text{d) } \sigma^2 = \lambda$$

$$\text{e) } \sigma^2 = \frac{1-p}{p^2}$$

$$\text{f) } \sigma^2 = \frac{Rm(M-R)(M-m)}{N^2(N-1)}$$

$$\text{g) } \sigma^2 = \left(\frac{v(1-p)}{p^2} \right)$$

Definition 4.22 Die Varianz einer diskret verteilten reellwertigen Zufallsgröße X mit der Verteilung P^X , gegeben durch $((x_i, p_i^X), i \in I)$, ist definiert als

$$\sigma_X^2 := E(X - EX)^2 = \sum_{i \in I} (x_i - EX)^2 p_i^X.$$

Man schreibt auch $\text{Var}(X)$ oder D^2X für σ_X^2 . Die Standardabweichung σ_X der Zufallsgröße X ist definiert als der Wert $(\sigma_X^2)^{\frac{1}{2}}$.

Offenbar gilt die für Berechnungen nützliche Formel

$$D^2X = EX^2 - (EX)^2$$

Aussage 4.23 (Tschebyschev'sche Ungleichung) Ist $0 < D^2X < \infty$, so gilt für jedes $\varepsilon > 0$ die Ungleichung

$$P(|X - EX| > \varepsilon) \leq \frac{D^2X}{\varepsilon^2}.$$

Beweis:

$$\begin{aligned} P(|X - EX| > \varepsilon) &= P^X(\{x_i : |x_i - EX| > \varepsilon\}) = \\ &= \sum_{\substack{i \in I \\ |x_i - EX| > \varepsilon}} P^X(\{x_i\}) \leq \sum_{i \in I} \frac{|x_i - EX|^2}{\varepsilon^2} P^X(\{x_i\}) = \frac{D^2X}{\varepsilon^2}. \end{aligned}$$

□

Die Tschebyschev'sche Ungleichung besagt, dass, je kleiner die Varianz von X ist, umso unwahrscheinlicher ist es, dass die Zufallsgröße X bei einer Durchführung des zugrunde liegenden zufälligen Versuches um mehr als ε vom Erwartungswert EX abweicht.

Im Fall $D^2X = 0$ gilt $P(X = EX) = 1$, es gibt also mit Wahrscheinlichkeit Eins keine Abweichung vom Erwartungswert, d.h. die Verteilung P^X ist ausgeartet und konzentriert in einem Punkt, der dann natürlich gleich EX ist.

Diskret verteilte zweidimensionale zufällige Vektoren

In vielen Fällen interessiert man sich im Rahmen eines zufälligen Versuches nicht nur für einzelne Zufallsgrößen, sondern für mehrere verschiedene. Diese sind dann im Allgemeinen nicht ohne innere Zusammenhänge, was man nur durch die Untersuchung ihrer gemeinsamen Wahrscheinlichkeitsverteilung feststellen kann und nicht an den einzelnen Zufallsgrößen bzw. ihren Verteilungen. Man denke beispielsweise an Körpergröße und Gewicht einer zufällig gewählten Person. Wir geben hier eine Einführung in diese Fragestellung im Rahmen zweier diskret verteilter Zufallsgrößen, sie bilden, zusammengefasst,

einen zweidimensionalen zufälligen Vektor.

Es sei $X = (U, V)^T$ ein zufälliger Vektor über einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Werten in einer Menge $E := E_U \times E_V$, wobei E_U und E_V höchstens abzählbar viele Elemente enthalten mögen:

$$E_U = \{u_i, i \in I\} \text{ und } E_V = \{v_j, j \in \mathfrak{J}\}.$$

Hier seien I und \mathfrak{J} Teilmengen von N_0 .

Die möglichen Werte der Zufallsgröße U sind also die $u_i \in E_U$, die von V die $v_j \in E_V$.

Die möglichen Werte von X sind die Paare $(u_i, v_j), (i, j) \in I \times \mathfrak{J}$. Folglich besitzt X eine diskrete Wahrscheinlichkeitsverteilung P^X . Ihre Einzelwahrscheinlichkeiten seien gegeben durch

$$P^X((u_i, v_j)) = P(U = u_i, V = v_j) =: p_{ij} \quad , \quad i \in I, j \in \mathfrak{J}.$$

Nach Definition diskreter Verteilungen gilt dann für die Wahrscheinlichkeit $P^X(B)$, dass der zufällige Vektor X einen Wert aus B annimmt (siehe Notation...):

$$P^X(B) = P(X \in B) = P((U, V) \in B) = \sum_{\substack{(i,j): \\ (u_i, v_j) \in B}} p_{ij} \quad , \quad B \subseteq E. \quad (4.18)$$

Definition 4.24 Die Verteilung P^X heißt gemeinsame Verteilung von U und V und ist gemäß Formel (4.20) eindeutig bestimmt durch ihre Einzelwahrscheinlichkeiten $p_{ij}, i \in I, j \in \mathfrak{J}$.

Die Verteilungen der einzelnen Zufallsgrößen U und V ergeben sich aus ihrer gemeinsamen Verteilung P^X durch

$$P^U(C) = P(U \in C) = P(U \in C, V \in E_V) = \sum_{\substack{i \in I: u_i \in C \\ j \in \mathfrak{J}}} p_{ij} \quad , \quad C \subseteq E_U \quad (4.19)$$

$$P^V(D) = P(V \in D) = P(U \in E_U, V \in D) = \sum_{\substack{j \in \mathfrak{J}: v_j \in D \\ i \in I}} p_{ij} \quad , \quad D \subseteq E_V \quad (4.20)$$

P^U und P^V sind also die Randverteilungen von P^X . Ihre Eigenschaften ergeben sich wie folgt:

$$P^U(\{u_i\}) = \sum_{j \in \mathfrak{J}} p_{ij} =: p_{i\cdot} \quad i \in I, P^V(\{v_j\}) = \sum_{i \in I} p_{ij} =: p_{\cdot j}, \quad j \in \mathfrak{J} \quad (4.21)$$

Die Bezeichnung Randverteilung wird hier besonders verständlich, wenn man die Einzelwahrscheinlichkeiten p_{ij} in einem Schema wie folgt anordnet.

$i \setminus j$	1	2	3	...	j	...	
1	p_{11}	p_{12}	p_{1j}	...	$p_{1\cdot}$
2	p_{21}	p_{22}					$p_{2\cdot}$
3	.						
.	.						
.	.						
.	.						
i	p_{i1}	p_{ij}	...
.	.						
	$p_{\cdot 1}$	$p_{\cdot 2}$				$p_{\cdot j}$	1

Bemerkung 4.25 Die Verteilung (p_{ij}) bestimmt die Randverteilungen $(p_{i\cdot})$ und $(p_{\cdot j})$ eindeutig. Die Randverteilungen bestimmen aber die gemeinsame Verteilung noch nicht eindeutig.

Das wird deutlich an dem nächsten Schema, das für jedes $c \in [0, \frac{1}{4}]$ eine zwei-dimensionale diskrete Verteilung darstellt:

	0	1	
0	$\frac{1}{4} + c$	$\frac{1}{4} - c$	$\frac{1}{2}$
1	$\frac{1}{4} - c$	$\frac{1}{4} + c$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Beispiel 4.26 Aus einem Kartenspiel mit 32 Karten (Skatspiel) werden nacheinander auf gut Glück ohne Zurücklegen der ersten Karte zwei Karten gezogen. Es sei $U = 1$ (bzw. $V = 1$), falls die erste (bzw. zweite) Karte ein König ist. Anderenfalls setzen wir $U = 0$ (bzw. $V = 0$). Dann ergibt sich unter Verwendung des Modells für die hypergeometrische Verteilung für die Einzelwahrscheinlichkeiten p_{ij} der gemeinsamen Verteilung von U und V und die Randverteilungen (vgl. den Abschnitt über hypergeometrische Verteilungen)

$U \setminus V$	0	1	
0	$\frac{28}{32} \cdot \frac{27}{31}$	$\frac{28}{32} \cdot \frac{4}{31}$	$\frac{7}{8}$
1	$\frac{4}{32} \cdot \frac{28}{31}$	$\frac{4}{32} \cdot \frac{3}{31}$	$\frac{1}{8}$
	$\frac{7}{8}$	$\frac{1}{8}$	

Funktionen diskret verteilter zufälliger Vektoren

Aussage 4.27 Es sei ψ eine reellwertige Funktion auf $E = E_U \times E_V$ mit Werten in einer höchstens abzählbar unendlichen Menge $E_W = \{w_k, k \in K\}$.

Dann ist

$$W(\omega) = \psi(U(\omega), V(\omega)), \quad \omega \in \Omega$$

eine diskret verteilte Zufallsgröße mit Werten in E_W und den Einzelwahrscheinlichkeiten

$$P^W(\{w_k\}) = \sum_{\substack{i,j: \\ \psi(u_i, v_j) = w_k}} p_{ij}, \quad k \in K. \quad (4.22)$$

Beweis:

$$P^W(\{w_k\}) = P(W = w_k) = P(\{\omega \in \Omega | W(\omega) = w_k\}) = P(W^{-1}(\{w_k\})) =$$

$$P(\{\omega : (U(\omega), V(\omega)) \in \psi^{-1}(\{w_k\})\}) = \sum_{\substack{(i,j): \\ \psi(u_i, v_j) = w_k}} p_{ij}, \quad k \in K.$$

□

Wir benötigen im Weiteren den Erwartungswert reellwertiger Funktionen mehrere Zufallsgrößen und nutzen dafür die folgende

Aussage 4.28 Gilt $E_W \subseteq$ und $\sum_{(i,j) \in I \times \mathfrak{J}} |\psi(u_i, v_j)| p_{ij} < \infty$, so hat $W = \psi(U, V)$ einen endlichen Erwartungswert, und es gilt

$$EW = E\psi(U, V) = \sum_{(i,j) \in I \times \mathfrak{J}} \psi(u_i, v_j) p_{ij} \quad (4.23)$$

Beweis:

$$\begin{aligned} EW &= \sum_{k \in K} w_k P(W = w_k) = \sum_{k \in K} w_k \sum_{\substack{(i,j): \\ \psi(u_i, v_j) = w_k}} p_{ij} = \\ &= \sum_{k \in K} \sum_{\substack{(i,j): \\ \psi(u_i, v_j) = w_k}} \psi(u_i, v_j) p_{ij} = \sum_{(i,j) \in I \times \mathfrak{J}} \psi(u_i, v_j) p_{ij}. \end{aligned}$$

□

Folgerungen 4.29 Sind U und V reellwertige Zufallsgrößen mit endlichem Erwartungswert und a, b reelle Zahlen, so hat auch $aU + bV$ einen endlichen

Erwartungswert, und es gilt

$$E(aU + bV) = aEU + bEV. \quad (4.24)$$

$$\text{Var}(aU + bV) = a^2\text{Var}(U) + b^2\text{Var}(V) + 2abE(u - EU)(V - EV) \quad (4.25)$$

Beweis: Wegen (4.25) gilt

$$\begin{aligned} E(aU + bV) &= \sum_{(i,j)} (au_i + bv_j)p_{ij} = a \sum_{i,j} u_i p_{ij} + b \sum_{i,j} v_j p_{ij} \\ &= a \sum_i u_i p_{i\cdot} + b \sum_j v_j p_{\cdot j} = aEU + bEV \end{aligned}$$

und

$$\begin{aligned} \text{Var}(aU + bV) &= E(aU + bV - E(aU + bV))^2 = E((aU - EaU) + (bV - EbV))^2 \\ &= a^2\text{Var}U + b^2\text{Var}V + 2abE(U - EU)(V - EV) \end{aligned}$$

□

Bemerkung 4.30 Im Allgemeinen gilt nicht $E(UV) = EUEV$. Das sieht man am Beispiel $\psi(U, V) = UV, P(U = i, V = j) = \frac{1}{4} + c \cdot (-1)^{i+j}, i, j \in \{0, 1\}$ für $c \in (0, \frac{1}{4})$.

4.4 Kovarianz und Korrelation

Es sei (U, V) ein diskret verteilter zufälliger Vektor über $(\Omega, \mathfrak{A}, P)$ mit Werten (u_i, v_j) in R_2 :

$$P(U = u_i, V = v_j) = p_{ij} \quad , \quad (i, j) \in I \times \mathfrak{J}$$

Aussage 4.31 (Cauchy-Schwarz-Ungleichung) Gilt $E(U^2) < \infty$ und $E(V^2) < \infty$, so ist $E|U \cdot V| < \infty$ und

$$(E(UV))^2 \leq EU^2EV^2. \quad (7) \quad (4.26)$$

Das Gleichheitszeichen gilt in (4.28) genau dann, wenn es eine reelle Zahl c gibt mit $U = cV$ P -f.s. oder mit $V = cU$ P -f.s.

(Eine Gleichung zwischen zwei Zufallsgrößen gilt P -fast sicher, kurz: P -f.s., falls die Menge aller $\omega \in \Omega$, für die sie nicht erfüllt ist, eine P -Nullmenge bildet.)

Beweis: O.B.d.A. sei $EU^2 > 0$ und $EV^2 > 0$. Anderenfalls gilt $U = 0$ P -f.s. oder $V = 0$ P -f.s.. Das Gleichheitszeichen in (4.28) und der zweite Teil der Aussage sind dann richtig.

Für jedes β aus R_1 ist $E(U + \beta V)^2 < \infty$ und zwar wegen $(a + b)^2 \leq 2(a^2 + b^2)$ ist $E(U + \beta V)^2 \leq 2EU^2 + 2\beta^2EV^2$ und der Voraussetzung.

Setzt man zunächst

$$\beta = \left(\frac{EU^2}{EV^2}\right)^{\frac{1}{2}} \quad \text{und dann} \quad \beta = -\left(\frac{EU^2}{EV^2}\right)^{\frac{1}{2}},$$

so erhält man wegen $E(U + \beta V)^2 \geq$ die Ungleichungen

$$-(EU^2EV^2)^{\frac{1}{2}} \leq E(UV) \leq (EU^2EV^2)^{\frac{1}{2}},$$

woraus sich (4.28) ergibt.

Das Gleichheitszeichen in (4.28) gilt wegen $EV^2 > 0$ genau dann, wenn $E(U + \beta V)^2 = 0$ für ein β aus R_1 richtig ist. In diesem Fall ist $U = -\beta V$ P -f.s. und notwendigerweise $\beta^2 = \frac{EU^2}{EV^2}$.

□

Definition 4.32 Es sei $E(U^2) < \infty$ und $E(V^2) < \infty$. Dann heißt die durch

$$Kov(U, V) := E((U - EU)(V - EV))$$

definierte Größe die Kovarianz zwischen U und V .

Aussage 4.33 Die Kovarianz hat folgende Eigenschaften (α, β seien zwei beliebige reelle Zahlen, W eine dritte Zufallsgröße):

1. $Kov(U, V) = Kov(V, U)$
2. $Kov(\alpha U, V) = \alpha Kov(U, V)$
3. $Kov(U + W, V) = Kov(U, V) + Kov(W, V)$
4. $Kov(U, V) = E(UV) - EUEV$
5. $Kov(U, U) = D^2U$
6. $Kov(U, \beta) = 0$
7. $(Kov(U, V))^2 \leq D^2U \cdot D^2V$
8. $(Kov(U, V))^2 = D^2UD^2V \iff \exists$ Es existieren $a, b \in R_1 : V = aU + b P - f.s.$ oder es existieren $c, d \in R_1 : U = cV + d P - f.s.$

Der Nachweis dieser Eigenschaften folgt für 1. - 6. unmittelbar aus der Definition der Kovarianz und für 7. und 8. mit Hilfe der Cauchy-Schwarz-Ungleichung.

Definition 4.34 Es sei $D^2U, D^2V \in (0, \infty)$. Dann bezeichnet man die Zahl

$$Kor(U, V) := \frac{Kov(U, V)}{(D^2UD^2V)^{\frac{1}{2}}}$$

als den Korrelationskoeffizienten zwischen U und V oder einfach als Korrelation zwischen U und V .

Wegen der Cauchy-Schwarz-Ungleichung gilt $|Kor(U, V)| \leq 1$.

Wir haben $|Kor(U, V)| = 1$ genau dann, wenn U und V linear (genauer: affin) abhängig sind, d. h., wenn es Zahlen a, b und c gibt mit $aU + bV + c = 0$ P -f.s. (Zum Beweis nutze man Eigenschaft 8 von Aussage 4.32.) Im letzteren Fall gilt $Kor(U, V) = 1$, falls $ab < 0$ und $Kor(U, V) = -1$ falls $ab > 0$.

Aussage 4.35 *Der Korrelationskoeffizient hat die Eigenschaften*

$$1.' \quad Kor(U, V) = Kor(V, U),$$

$$2.' \quad Kor(\alpha U, V) = Kor(U, V).$$

Definition 4.36 *Gilt für zwei Zufallsgrößen U, V mit $D^2U < \infty$ und $D^2V < \infty$ die Beziehung $Kor(U, V) = 0$, so heißen U und V unkorreliert.*

Die Größe $Kor(U, V)$ gibt den Grad der linearen Abhängigkeit zwischen den Zufallsgrößen U und V an. Für $Kor(U, V) = 1$ und $Kor(U, V) = -1$ liegt vollständige lineare Abhängigkeit vor. $Kor(U, V) = 0$ deutet auf eine gewisse Unabhängigkeit in einem noch zu präzisierenden Sinn.

Man beachte, dass auf Grund der Definition der Eigenschaft 4. der Aussage 4.32 gilt

$$Kor(U, V) = 0 \iff Kov(U, V) = 0 \iff E(UV) = EU \cdot EV \quad (4.27)$$

4.5 Regressionsgerade

Wir beginnen mit einer Vorüberlegung über die beste Möglichkeit, den Wert einer Zufallsgröße, den sie bei dem ihr zugrunde liegenden zufälligen Versuch annehmen wird, vorherzusagen.

Es sei X eine reellwertige (diskret verteilte) Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit $D^2X < \infty$.

Wenn man vor Ausführung des zufälligen Versuches $(\Omega, \mathfrak{A}, P)$ den Wert, den X annehmen wird, durch eine reelle Zahl c voraussagen soll, so ist das im Fall $D^2X > 0$ zunächst einmal nicht mit Sicherheit möglich. Um es dennoch so gut wie möglich zu tun, muss man präzisieren, was man unter "so gut wie möglich" verstehen will. Eine Möglichkeit besteht darin, zu fordern, dass in einer langen Reihe von Realisierungen von X , nämlich (x_1, x_2, \dots, x_n) , c so gewählt wird, dass $\sum_{i=1}^n (x_i - c)^2$ minimal wird ("Minimierung der quadratischen Abweichung", "Methode der kleinsten Quadrate").

Das führt auf $c = \frac{1}{n} \sum_{i=1}^n x_i$. Das empirische Gesetz der großen Zahlen besagt, dass dieses arithmetische Mittel für die Zufallsgröße X , in der Nähe von EX liegt.

Wir machen uns von der Durchführung des Versuches unabhängig und verwenden als Vorhersage von X den Wert $c = EX$. Tatsächlich erreicht auch die Funktion $c \rightarrow E(X - c)^2$ bei $c = EX$ ein Minimum. Die "beste" Voraussage für X ist also EX (im Sinne der Minimierung des quadratischen Mittels).

Die Streuung $D^2X = E(X - EX)^2$ ist gerade der Wert dieses Minimums und bildet ein Maß für die "Güte" der Voraussage von X durch EX . Je kleiner D^2X ist, umso genauer ("im quadratischen Mittel") wird diese Voraussage sein.

Wir wenden uns nun dem eigentlichen Anliegen dieses Abschnittes zu.

Es seien U und V zwei (diskret verteilte) reellwertige Zufallsgrößen über demselben Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit $0 < EU^2 < \infty, 0 < EV^2 < \infty$. Die Aufgabe bestehe darin, auf Grundlage der Kenntnis, welchen Wert U angenommen hat, den Wert von V möglichst gut vorherzusagen. Zur Illustration stelle man sich wieder den Fall vor, dass U die Körpergröße und V das Gewicht einer zufällig ausgewählten Person sind.

Im Allgemeinen gibt es keine deterministische Funktion ψ , so dass $V = \psi(U)$ gilt. Um V mit Hilfe von U möglichst gut voraussagen, suchen wir Koeffizienten $a, b \in R_1$, die die mittlere quadratische Abweichung

$$(a, b) \longrightarrow E(V - aU - b)^2 =$$

$$EV^2 + a^2EU^2 + b^2 - 2aE(UV) - 2bEV + 2abEU$$

minimal werden lassen, d. h., wir suchen unter allen linearen Funktionen von U diejenige, die V am besten approximiert.

Das führt auf die Gleichungen

$$b = EV - aEU \text{ und}$$

$$aD^2U = Kov(U, V).$$

Also ist

$$\hat{V} := EV + Kor(U, V) \cdot \frac{\sigma_V}{\sigma_U}(U - EU)$$

die beste lineare Approximation von V durch U . **Definition:** Die Gerade $v = g(u) = EV + a(u - EU)$, $u \in R_1$

$$\text{mit } a = Kor(U, V) \left(\frac{\sigma_V}{\sigma_U} \right) = \frac{Kov(U, V)}{\sigma_U^2}$$

heißt *Regressionsgerade* für V bezüglich U . Die Zufallsgröße $\hat{V} = g(U)$ ist die (im quadratischen Mittel) beste lineare Funktion von U für die Voraussage von V auf der Basis von U (= *Regressionsgerade* für V auf der Basis von U).

Man wird mit der Vorhersage \hat{V} für V den tatsächlich eintretenden Wert von V i. Allg. nicht genau treffen. Im Mittel allerdings schon, denn es gilt $E\hat{V} = EV$. Die tatsächliche "Ungenauigkeit" $\hat{V} - V$ hängt vom Zufall ab. Wir messen sie durch ihre Varianz $E(\hat{V} - V)^2$, für die sich nach einfacher Rechnung

$$E(V - \hat{V})^2 = \sigma_V^2(1 - (Kor(U, V))^2)$$

ergibt. Diese Zahl bezeichnet man als *Reststreuung*, die zwischen der Vorhersage \hat{V} und dem vorherzusagendem Wert V noch besteht, und die man auf Grundlage der Vorhersage von V durch eine lineare Funktion von U nicht beseitigen kann.

Hier wird noch einmal deutlich, dass $Kor(U, V)$ ein Maß für den linearen Zusammenhang zwischen U und V ist.

Spezialfälle:

- a) $Kor(U, V) = 0 \implies$ keine Reduzierung von σ_V^2 , die beste lineare Funktion \hat{V} zur Vorhersage von V auf der Basis von U hängt gar nicht von U ab und ist gleich dem Wert EV .
- b) $|Kor(U, V)| = 1 : \hat{V} = V$, keine Reststreuung, exakte Voraussage von V durch eine lineare Funktion von U möglich

4.6 Erzeugende Funktionen

Für diskrete Verteilungen auf den natürlichen Zahlen stellen die sogenannten *erzeugenden Funktionen* ein wirkungsvolles analytisches Hilfsmittel dar, um zum Beispiel Momente der Verteilung zu bestimmen. Weitere Anwendungen werden wir später kennen lernen.

Es sei X eine Zufallsgröße über einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$, die nur Werte aus der Menge N_0 der natürlichen Zahlen annehmen kann, und mit Einzelwahrscheinlichkeiten ihrer Verteilung

$$p_k = P(X = k), \quad k \geq 0.$$

Definition 4.37 Als erzeugende Funktion $g(s)$, $s \in [-1, 1]$, der Zufallsgröße X (genauer: ihrer Wahrscheinlichkeitsverteilung P^X) bezeichnet man die Funktion

$$g(s) := E(s^X) = \sum_{k \geq 0} s^k p_k, \quad s \in [-1, 1].$$

Wegen $p_k \geq 0$ und $\sum_{k \geq 0} p_k = 1$ ist $g(\cdot)$ eine Potenzreihe mit einem Konvergenzradius $\rho \geq 1$. Daraus ergeben sich sofort einige Eigenschaften, die wir in folgender Aussage zusammenfassen.

Aussage 4.38 In der soeben eingeführten Terminologie gilt

(i) $g(\cdot)$ ist in $(-1, 1)$ unendlich oft differenzierbar mit

$$\frac{d^k}{ds^k} g(s) = \sum_{j \geq k} j(j-1) \cdots (j-k+1) s^{j-k} p_j =$$

$$E(X(X-1) \cdots (X-k+1) s^{X-k}),$$

es gilt

$$p_k = \frac{1}{k!} \frac{d^k}{ds^k} g(s) \Big|_{s=0}, \quad k \geq 0. \quad (4.28)$$

(ii) Im Fall $EX^k < \infty$ haben wir die Gleichung

$$E(X(X-1)(X-2) \cdots (X-k+1)) = \lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) < \infty. \quad (4.29)$$

Gilt dagegen $EX^k = \infty$, so ist

$$E(X(X-1) \cdots (X-k+1)) = \lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) = \infty.$$

(iii) Sind $g(\cdot)$ und $h(\cdot)$ erzeugende Funktion zweier Zufallsgrößen X bzw. Y mit Werten in N_0 , und gilt $g(s) = h(s)$, $s \in [0, \delta]$, für ein $\delta > 0$, so sind die Verteilungen P^X und P^Y einander gleich:

$$P(X = k) = P(Y = k), \quad k \geq 0.$$

Beweis:

- (i) Es sei $|s| < 1$ und $\delta > 0$, so dass $(s - \delta, s + \delta) \subseteq (-1, 1)$.
Dann ist für alle $h \in (-\delta, \delta)$

$$A(s, h) := |h^{-1}[g(s+h) - g(s)] - \sum_{k=1}^{\infty} k s^{k-1} p_k| =$$

$$\sum_{k \geq 1} [h^{-1}((s+h)^k - s^k) - k s^{k-1}] p_k$$

Weiterhin gibt es für jedes $k \geq 2$ ein ξ_k mit $|\xi_k| \leq h$, so dass gilt

$$h^{-1}((s+h)^k - s^k) - k s^{k-1} = \frac{k(k-1)}{2} \cdot (s + \xi_k)^{k-2} \cdot h$$

(Mittelsatzwert). Wegen $|s + \xi_k| \leq |s| + \delta < 1$ ergibt sich

$$|A(s, h)| \leq \sum_{k \geq 1} \frac{h(k-1)}{2} (|s| + \delta)^k p_k \cdot |h| = 0(h).$$

Für $h \rightarrow 0$ folgt also

$$\frac{dg}{ds} = \sum_{k \geq 1} k s^{k-1} p_k, \text{ und es gilt } \frac{dg}{ds} \Big|_{s=0} = p_1.$$

Der Beweis für die höheren Ableitungen erfolgt analog.

- (ii) Mit $EX^k < \infty$ gilt auch $EX^l < \infty$ ($1 \leq l < k$) und somit $E(X(X-1)\dots(X-k+1)) < \infty$. Für $s \in (0, 1)$ ist $\frac{d^k}{ds^k} g(s)$ eine nichtnegative monoton wachsende Funktion mit (siehe Teil (i) dieser Aussage)

$$\lim_{s \uparrow 1} \frac{d^k}{ds^k} g(s) \leq E(X(X-1)\dots(X-k+1)) < \infty. \quad (4.30)$$

Es sei ε irgendeine positive Zahl und j_0 so groß, dass

$$\sum_{j=j_0+1}^{\infty} j(j-1)\dots(j-k+1)p_j < \frac{\varepsilon}{2}$$

gilt.

Weiterhin sei $\delta < 0$ so gewählt, dass

$$\sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)s^j p_j > \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)p_j - \frac{\varepsilon}{2}$$

für alle s mit $s \in (1 - \delta, 1]$ richtig ist.

Dann gilt für $s \in (1 - \delta, 1]$

$$\begin{aligned} \frac{d^k g}{s d^k}(s) &= \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)s^j p_j + \sum_{j=j_0+1}^{\infty} j(j-1)\dots(j-k+1)s p_j \\ &> \sum_{j=k}^{j_0} j(j-1)\dots(j-k+1)p_j - \frac{\varepsilon}{2} > \sum_{j=k}^{\infty} j(j-1)\dots(j-k+1)p_j - \varepsilon, \end{aligned}$$

und somit haben wir in (4...) das Gleichheitszeichen.

(iii) Nach Voraussetzung und wegen (i) gilt

$$\frac{d^k g}{ds^k}(s) = \frac{d^k h}{ds^k}(s), \quad k \geq 1, s \in (0, \delta).$$

Wegen der Stetigkeit aller Ableitungen von g und von h für $|s| < 1$ folgt

$$\left. \frac{d^k g}{ds^k} \right|_{s=0} = \left. \frac{d^k h}{ds^k} \right|_{s=0}.$$

Aus (4...) ergibt sich nun (iii). □

Definition 4.39 Die Größe $f_k := EX(X-1)\dots(X-k+1)$ heißt faktorielles Moment k -ter Ordnung der Zufallsgröße X .

Formel (4...) kann man zur Berechnung anderer Momente der Zufallsgröße X nutzen. Zum Beispiel gilt

$$EX = f_1, D^2 X = EX^2 - (EX)^2 = f_2 + f_1 - f_1^2.$$

Beispiel 4.40 (Fortsetzung der Beispiele aus 4.1.):

a) Im Fall $\omega_k = k, k = 1, \dots, N$ ergibt sich

$$g(s) = \frac{1}{N} \sum_{k=1}^N s^k = \frac{1}{N} \cdot \frac{s - s^{N+1}}{1 - s}, \quad s \in [-1, 1)$$

und $g(1) = 1$,

b) $g(s) = s^{k_0}$ falls $\omega_0 = k_0 \in N_0$,

$$c) g(s) = \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} = (1 - p(1-s))^n,$$

$$d) g(s) = \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} e^{-\lambda} = \exp(\lambda(s-1)),$$

$$e) g(s) = \sum_{k=0}^{\infty} (qs)^k p = \frac{p}{1-qs} \text{ mit } q = 1 - p,$$

$$f) g(s) = \sum_{k=0}^m \frac{\binom{R}{k} \binom{M-R}{m-k}}{\binom{M}{m}} s^k \text{ ist eine spezielle hypergeometrische Funktion,}$$

$$g) g(s) = \sum_{k=0}^{\infty} \binom{-v}{k} (qs)^k p = \left(\frac{p}{1-qs}\right)^v \text{ mit } q = 1 - p.$$

Der Beweis ist elementar.

4.7 Mehrstufige zufällige Versuche

Häufig läuft ein zufälliger Versuch in mehreren Schritten oder Stufen ab. Wir haben dafür bereits Beispiele kennen gelernt (mehrmaliges Werfen einer Münze). In diesem Abschnitt werden wir zunächst ein sehr allgemeines stochastisches Modell zusammengesetzter Versuche konstruieren. Danach konzentrieren wir uns auf den Fall abzählbar vieler Versuchsausgänge, in dem man

einige einfache Berechnungsformeln angeben kann.

Angenommen, der zufällige Versuch besteht aus n Einzelexperimenten, die nacheinander ausgeführt werden. Die möglichen Ergebnisse ω_k des k -ten Experimentes mögen aus einer Menge Ω_k stammen, $k = 1, \dots, n$. Das Ergebnis des Gesamtexperimentes wird dann beschrieben durch den Ausgang

$$\omega = (\omega_1, \dots, \omega_n) \in \Omega_1 \times \dots \times \Omega_n.$$

Da ω aufgefasst wird als Ergebnis einer zeitlichen Abfolge von Experimenten, nennt man ω auch einen "Pfad" oder eine "Trajektorie" des Gesamtversuches.

Wir setzen

$$\Omega := \Omega_1 \times \dots \times \Omega_n = \prod_{k=1}^{n \otimes} \Omega_k.$$

Die mit dem k -ten Experiment verbundenen Ereignisse bilden eine σ -Algebra \mathfrak{A}_k von Teilmengen von Ω_k . Die σ -Algebra \mathfrak{A} aller mit dem Gesamtversuch verbundenen Ereignisse enthält natürlich alle Ereignisse der Form $A := A_1 \times \dots \times A_n$ mit $A_k \in \mathfrak{A}_k$, $k = 1, \dots, n$, da man nach Ablauf aller Telexperimente entscheiden kann, ob ein solches A eingetreten ist oder nicht.

Wir definieren \mathfrak{A} als kleinste σ -Algebra von Teilmengen von Ω , die alle Ereignisse dieser Form umfasst, also:

$$\mathfrak{A} := \sigma(A_1 \times \dots \times A_n | A_k \in \mathfrak{A}_k, k = 1, \dots, n).$$

Definition 4.41 \mathfrak{A} heißt die Produkt- σ -Algebra der σ -Algebren \mathfrak{A}_k , $k = 1, \dots, n$, und wird auch mit $\prod_{k=1}^{n \otimes} \mathfrak{A}_k$ oder $\mathfrak{A}_1 \otimes \mathfrak{A}_2 \otimes \dots \otimes \mathfrak{A}_n$ bezeichnet.

Ist P ein Wahrscheinlichkeitsmaß auf \mathfrak{A} , so haben wir mit $(\Omega, \mathfrak{A}, P)$ ein stochastisches Modell eines n -stufigen zufälligen Versuches.

Das System γ von Ereignissen aus \mathfrak{A} , definiert durch

$$\gamma := \{A_1 \times \dots \times A_n | A_k \in \mathfrak{A}_k, k = 1, \dots, n\}$$

ist eine Semialgebra mit $\sigma(\gamma) = \mathfrak{A}$. Folglich ist P durch die Angabe seiner Werte auf γ bereits eindeutig festgelegt (Maßtheorie). Das wird uns die Konstruktion des Maßes P aus einfacheren Größen ermöglichen.

Das k -te Einzelexperiment $(\Omega_k, \mathfrak{A}_k, P_k)$ ist in dem Gesamtexperiment $(\Omega, \mathfrak{A}, P)$ durch

$$\mathfrak{A}_k \ni A_k \longleftrightarrow (\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n) =: A'_k \text{ und}$$

$$P_k(A_k) = P(A'_k), \quad A_k \in \mathfrak{A}_k$$

eingebettet. Die Verteilung P bestimmt also die "Randverteilungen" P_k auf \mathfrak{A}_k .

Aus den $P_k, k = 1, \dots, n$, dagegen ist P im Allgemeinen nicht reproduzierbar. Das gelingt nur in einem Fall, nämlich wenn gilt

$$P(A_1 \times \dots \times A_n) = \prod_{k=1}^n P(A_k), \quad A_k \in \mathfrak{A}_k, k = 1, \dots, n. \quad (4.31)$$

In diesem Fall bezeichnet man P als das von den P_k erzeugte *Produktmaß* auf der Produkt- σ -Algebra \mathfrak{A} und schreibt $P = \prod_{k=1}^n P_k = P_1 \otimes P_2 \otimes \dots \otimes P_n$.

Im Allgemeinen ist jedoch P nicht gleich dem Produktionsmaß.

Wir wollen nun für den Fall, dass alle Ω_k abzählbar sind, das Maß P aus einfacheren Kenngrößen konstruieren. Dazu beginnen wir mit einem einfachen Beispiel.

Beispiel 4.42 In einer Urne mögen sich zwei rote und drei schwarze Kugeln befinden. Wir ziehen auf gut Glück eine der Kugeln und legen sie zusammen mit einer weiteren Kugel derselben Farbe wie die gezogene, in die Urne zurück. Danach wählen wir erneut auf gut Glück eine Kugel.

Das Experiment ist zweistufig mit $\Omega_1 = \Omega_2 = \{r, s\}$, seine möglichen Ausgänge sind die Elemente der Menge $\Omega := \{(r, r), (r, s), (s, r), (s, s)\}$. Für \mathfrak{A} wählen wir $\mathfrak{P}(\Omega)$. Die zu bestimmende Wahrscheinlichkeitsverteilung P ist diskret und durch ihre Einzelwahrscheinlichkeiten $p((r, r)), p((r, s)), p((s, r)), p((s, s))$ eindeutig festgelegt.

$$P(A) = \sum_{\omega: \omega \in A} p(\omega), \quad A \subseteq \Omega. \quad (4.32)$$

Um eine Vorstellung zu bekommen, wie groß diese Einzelwahrscheinlichkeiten im betrachteten Fall sind, erinnern wir an das empirische Gesetz der großen Zahlen, dass bei wachsender Zahl von Versuchsdurchführungen die relative Häufigkeit $\frac{n(A)}{n}$ eines Ereignisses A sich der Wahrscheinlichkeit $P(A)$ immer mehr nähert. Wenn wir die geschilderten Ziehungen sehr oft wiederholen, so wird die relative Häufigkeit, beim jeweils ersten Zug eine rote Kugel zu erhalten, etwa gleich $\frac{2}{5}$ sein, da in der Urne zwei der fünf Kugeln rot sind. Unter denjenigen Versuchsdurchführungen, bei denen man beim ersten Mal eine rote Kugel zieht, werden sich mit der relativen Häufigkeit von etwa $\frac{3}{6} = \frac{1}{2}$ beim zweiten Ziehen eine schwarze Kugel ergeben, da sich vor dem zweiten Ziehen drei rote und drei schwarze Kugeln in der Urne befinden. Insgesamt wird also die relative Häufigkeit des Ergebnisses (r, s) etwa gleich $\frac{2}{5} \cdot \frac{1}{2} = \frac{1}{5}$ sein. Wir setzen deshalb die Einzelwahrscheinlichkeit $p((r, s))$ dafür, beim ersten Zug eine rote, beim zweiten Zug eine schwarze Kugel zu erhalten, gleich $p((r, s)) = \frac{2}{5} \cdot \frac{1}{2} = \frac{1}{5}$. Analog ergibt sich $p((r, r)) = \frac{2}{5} \cdot \frac{3}{6} = \frac{1}{5}$, $p((s, r)) = \frac{3}{5} \cdot \frac{2}{6} = \frac{1}{5}$, $p((s, s)) = \frac{3}{5} \cdot \frac{4}{6} = \frac{2}{5}$. Damit ist unter Beachtung von (4...) eine Verteilung auf $\mathfrak{B}(\Omega)$ definiert.

Für die Randverteilungen P_1 und P_2 des ersten bzw. zweiten Zuges ergibt sich

$$P_1(\{r\}) = P(\{r\} \times \{r, s\}) = p((r, r)) + p((r, s)) = \frac{2}{5},$$

$$P_1(\{s\}) = 1 - P_1(\{r\}) = \frac{3}{5}$$

$$P_2(\{r\}) = P(\{r, s\} \times \{r\}) = p((r, r)) + p((s, r)) = \frac{2}{5}, \quad P_2(\{s\}) = 1 - P_2(\{r\}) = \frac{3}{5}.$$

Erste Pfadregel

Im Folgenden seien alle $\Omega_k, k = 1, \dots, n$, höchstens abzählbar. Das erste der n Experimente ende mit der Wahrscheinlichkeit $p^{(1)}(\omega_1)$ mit dem Ausgang $\omega_1 \in \Omega_1$. Die Wahrscheinlichkeitsverteilung für das zweite Experiment hängt

i.A. vom Ausgang des ersten ab, wir bezeichnen ihre Einzelwahrscheinlichkeiten mit $p_{\omega_1}^{(2)}(\omega_2)$. Analog hängt die Wahrscheinlichkeitsverteilung des k -ten Experiments vom bisherigen Verlauf ab: $p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k)$. Als Wahrscheinlichkeit $p(\omega)$ für den Ausgang des Gesamtexperimentes $\omega = (\omega_1, \dots, \omega_n)$ definieren wir analog zum obigen Beispiel:

$$p(\omega) := p^{(1)}(\omega_1)p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot p_{\omega_1\dots\omega_{n-1}}^{(n)}(\omega_n). \quad (4.33)$$

Offenbar ist $p(\omega) \geq 0$ und

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{\omega_1 \in \Omega_1} p^{(1)}(\omega_1) \sum_{\omega_2 \in \Omega_2} p_{\omega_1}^{(2)}(\omega_2) \cdots \sum_{\omega_n \in \Omega_n} p_{\omega_1\dots\omega_{n-1}}^{(n)}(\omega_n) = 1,$$

da nach Definition gilt:

$$p^{(1)}(\omega_1) \geq 0 \text{ und } \sum_{\omega_1 \in \Omega_1} p^{(1)}(\omega_1) = 1, \text{ sowie}$$

$$p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k) \geq 0 \text{ und}$$

$$\sum_{\omega_k \in \Omega_k} p_{\omega_1\omega_2\dots\omega_{k-1}}^{(k)}(\omega_k) = 1, \quad k \geq 2.$$

Diese Regel (4....) zur Bestimmung der Einzelwahrscheinlichkeiten $p(\omega)$ heißt auch "*Erste Pfadregel*".

Die Wahrscheinlichkeit $p(\omega)$ des Pfades $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ wird also mittels der zumeist bekannten oder einfacher zu bestimmenden Größen $p_{\omega_1\dots\omega_{k-1}}^{(k)}(\omega_k)$ berechnet, die die Wahrscheinlichkeiten angeben, dass beim k -ten Versuch der Ausgang ω_k erscheint, wenn im bisherigen Verlauf der Versuche $\omega_1, \omega_2, \dots, \omega_{k-1}$ aufgetreten sind.

Zweite Pfadregel

Als "*Zweite Pfadregel*" bezeichnet man die Formel zur Bestimmung von $P(A)$, wie sie für jede diskrete Verteilung gilt:

$$P(A) = \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega. \quad (4.34)$$

Als Fazit halten wir fest, dass die Wahrscheinlichkeitsverteilung P in einem n -stufigen zufälligen Versuch sich aus den meist einfacher zu bestimmenden Wahrscheinlichkeiten

$$p^{(1)}(\omega_1), p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k), k = 2, 3, \dots, n, \omega_i \in \Omega_i, i = 1, \dots, n$$

mit Hilfe der ersten und zweiten Pfadregel bestimmen lässt. Die Verteilung $(p^{(1)}(\omega_1), \omega_1 \in \Omega_1)$ nennt man *Anfangsverteilung*, jede der Verteilungen $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k), \omega_k \in \Omega_k$, eine *Übergangsverteilung*.

Beispiel 4.43 (Polya'sches Urnenschema)

In einer Urne liegen R rote und S schwarze Kugeln, $R + S = N$. Auf rein zufällige Weise wird n -mal nacheinander eine Kugel entnommen, jedes Mal wieder zurückgelegt und c Kugeln derselben Farbe hinzu gefügt.

Die Anzahl der roten Kugeln in der Urne nach Abschluss den Ziehungen und des Zurücklegens ist eine Zufallsgröße. Gefragt ist ihre Wahrscheinlichkeitsverteilung. Der beschriebene Mechanismus wird als einfaches Modell der Ausbreitung infektiöser Krankheiten in Populationen von Individuen angesehen und heißt *Polya'sches Urnenschema*.

Wir beschreiben das Polya'sche Urnenschema als mehrstufigen zufälligen Versuch. Dazu setzen wir

$$\Omega_k = \{0, 1\}, \mathfrak{A}_k = \mathfrak{P}(\Omega_k), k = 1, \dots, n$$

und vereinbaren $\omega_k = 0$, falls beim k -ten Ziehen eine schwarze, $\omega_k = 1$, falls beim k -ten Ziehen eine rote Kugel erscheint.

Es sei $\Omega := \prod_{k=1}^n \Omega_k$, $\mathfrak{A} := \mathfrak{P}(\Omega)$, $\omega = (\omega_1, \dots, \omega_n) \in \Omega$.

Die Anzahl der roten Kugeln, die bis zum (einschließlich) k -ten Ziehen gezogen wurden, werde mit R_k bezeichnet.

Die Zufallsgrößen R_k und S_k habe die möglichen Werte $0, 1, \dots, k$. Nach Defi-

inition gilt $R_k(\omega) = \sum_{j=1}^k \omega_j$, $\omega \in \Omega, k = 1, \dots, n$.

Wir setzen $R_0(\omega) \equiv 0$.

Die Anzahl S_k der schwarzen Kugeln, die bis zum k -ten Ziehen erschienen sind, ist folglich gleich $k - R_k(\omega) =: S_k(\omega)$. Auch hier definieren wir $S_0(\omega) \equiv 0$.

Nach den beiden Pfadregeln für mehrstufige Versuche gilt:

$$(R_n = j) = \sum_{\substack{\omega \in \Omega \\ R_n(\omega) = j}} P(\{\omega\})$$

$$P(\{\omega\}) = p^{(1)}(\omega_1) p_{\omega_1}^{(2)}(\omega_2) \cdot \dots \cdot p_{\omega_1 \dots \omega_{n-1}}^{(n)}(\omega_n) \quad (9) \quad (4.35)$$

Die Übergangswahrscheinlichkeit $p_{\omega_1 \dots \omega_{k-1}}^{(k)}(\omega_k)$ ist dabei die Wahrscheinlichkeit dafür, dass im k -ten Versuch eine rote ($\omega_k = 1$) bzw. eine schwarze ($\omega_k = 0$) Kugel gezogen wird, wobei $(\omega_1, \dots, \omega_{k-1})$ den bisherigen Verlauf der Ziehungen darstellt.

Lemma 4.44 *Es gilt*

$$p_{\omega_k \dots \omega_{k-1}}^{(k)}(\omega_k) = \frac{(R + R_{k-1}(\omega) \cdot c)^{\omega_k} (S + S_{k-1}(\omega) \cdot c)^{1-\omega_k}}{R + S + (k-1)c}, \quad k = 1, \dots, n. \quad (4.36)$$

Beweis: Vor dem k -ten Ziehen einer Kugel befinden sich $R + S + (k-1)c$ Kugeln in der Urne, von denen $R + R_{k-1}(\omega)c$ rot und $S + S_{k-1}(\omega) \cdot c$ schwarz sind. \square

Beim Einsetzen von (4.36) in (4.35) kann man das entstehende Produkt vereinfachen. Dazu nutzen wir das folgende

Lemma 4.45 *Für alle ω aus Ω gilt*

$$\prod_{k=0}^{n-1} (R + R_k(\omega)c)^{\omega_{k+1}} = \prod_{k=1}^{R_n(\omega)} (R + (k-1)c) \quad \text{und} \quad (11) \quad (4.37)$$

$$\prod_{k=0}^{n-1} (S + S_k(\omega)c)^{1-\omega_{k+1}} = \prod_{k=1}^{S_n(\omega)} (S + (k-1)c), \quad (4.38)$$

wobei $\prod_{k=1}^0 (\dots) = 1$ gesetzt wird.

Beweis: Wir zeigen nur, dass die erste Gleichung richtig ist, die zweite beweist man analog.

Für $n = 1$ ist die Gleichheit in (11) offensichtlich. Die Beziehung (11) gelte für $n = m$. Dann haben wir für $n = m + 1$:

$$\begin{aligned} \prod_{k=0}^m (R + R_k(\omega)c)^{\omega_{k+1}} &= \prod_{k=0}^{m-1} (R + R_k(\omega)c)^{\omega_{k+1}} \cdot (R + R_m(\omega)c)^{\omega_{m+1}} \\ &= \prod_{k=1}^{R_m(\omega)} (R + (k-1)c) (R + R_m(\omega)c)^{\omega_{m+1}}. \end{aligned}$$

Der letzte Faktor ist gleich $(R + (R_{m+1}(\omega) - 1)c)$ falls $\omega_{m+1} = 1$ (wegen $R_{m+1}(\omega) = R_m(\omega) + \omega_{m+1}$), und gleich Eins, falls $\omega_{m+1} = 0$. Wegen $R_{m+1}(\omega) = R_m(\omega)$ im letzteren Fall ist somit (11) für $n = m + 1$ und jedes ω richtig. \square

Folgerungen 4.46 Für jedes $\omega \in \Omega$ gilt

$$P(\{\omega\}) = \frac{\prod_{j=1}^{R_n(\omega)} (R + (j-1)c) \prod_{j=1}^{n-R_n(\omega)} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)} \quad (4.39)$$

$$P(R_n = l) = \sum_{\omega: R_n(\omega)=l} P(\{\omega\}), l = 0, 1, \dots, n.$$

Alle ω mit $R_n(\omega) = l$ haben wegen (13) die gleiche Wahrscheinlichkeit $P(\{\omega\})$. Es gibt insgesamt $\binom{n}{l}$ von ihnen (das ist die Anzahl aller möglichen Anordnungen von l Einsen und $(n-l)$ Nullen in $\omega = (\omega_1, \dots, \omega_n)$). Also folgt

$$P(R_n = l) = \frac{\binom{n}{l} \prod_{j=1}^l (R + (j-1)c) \prod_{j=1}^{n-l} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)} \quad (4.40)$$

$$l = 0, 1, \dots, n.$$

Diese Wahrscheinlichkeitsverteilung von R_n heißt "Polya'sche Verteilung" auf $\{0, 1, \dots, n\}$ mit den Parametern R, S und c .

Damit haben wir:

Aussage 4.47 Die Anzahl der bis zum n -ten Zug gezogenen roten Kugeln im Polya'schen Urnenschema hat die durch (14) gegebene Polya'sche Verteilung mit den Parametern R, S und c .

Spezialfälle:

$c = 0$: In diesem Fall wird nach jedem Ziehen nur die gezogene Kugel selbst zurückgelegt. Es ergibt sich eine Binomialverteilung mit den Parametern n und $p = \frac{R}{R+S}$,

$c = -1$: Jetzt wird nach jedem Ziehen die gezogene Kugel nicht zurückgelegt und auch keine weitere in die Urne gelegt. Wir erhalten eine hypergeometrische Verteilung mit den Parametern $M = R + S, m = n$. (Wir setzen in diesem Fall der Einfachheit halber $m \leq \min(R, S)$ voraus.)

Es sei $A_k =$ "bei k -ten Zug erscheint eine rote Kugel", $1 \leq k \leq n$. Wir wollen $P(A_k)$ berechnen und beweisen zunächst das folgende Lemma.

Lemma 4.48 Die Ereignisse A_1, A_2, \dots, A_n sind austauschbar im Sinne von:

Beweis: Wegen (4....) gilt für alle Tupel (i_1, \dots, i_l) mit $1 \leq i_1 < \dots < i_l \leq n$ die Gleichung

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l}) = \sum_{\substack{\omega \in \Omega \\ \omega_{i_j} = 1, j=1, \dots, l}} P(\{\omega\}) = \sum_{\substack{\omega \in \Omega: \\ \omega_{i_j} = 1, j=1, \dots, l}} H_n(R_n(\omega)) = \sum_{m=l}^n \sum_{\substack{\omega \in \Omega \\ \omega_{i_j} = 1, j=1, \dots, l \\ R_n(\omega) = m}} H_n(m) \quad (4.41)$$

mit

$$H_n(m) = \frac{\prod_{j=1}^m (R + (j-1)c) \prod_{j=1}^{n-m} (S + (j-1)c)}{\prod_{j=1}^n (R + S + (j-1)c)}, \quad 0 \leq m \leq n.$$

Die Summanden $H_n(m)$ der rechten Seite von (4.41) sind für alle Tupel (i_1, \dots, i_l) mit derselben Länge l dieselben, und die Anzahl der ω mit $\omega_{i_k} = 1, k = 1, \dots, l$, und $R_n(\omega) = m$ ist bei m mit $l \leq m \leq n$ gleich $\binom{n-l}{m-l}$, unabhängig von den Werten $i_1, i_2, \dots, i_l \in \{1, 2, \dots, n\}$. Somit gilt $P(A_{i_1} \cap \dots \cap A_{i_l}) = P(A_1 \cap \dots \cap A_l)$, die $(A_k, k \leq n)$ sind also austauschbar im Sinne von

Mit Hilfe dieses Lemmas kommt man zu der zunächst überraschenden

$$P(A_k) = P(A_1) = \frac{R}{R+S}, \quad k = 1, 2, \dots, n$$

unabhängig von k und c .

Aussage 4.49 Für alle $k = 1, 2, \dots, n$ gilt

$$P(A_k) = P(A_1) = \frac{R}{R+S}$$

Beweis: Wegen der Austauschbarkeit der A_1, A_2, \dots, A_n folgt insbesondere $P(A_k) \equiv P(A_1)$, und $P(A_1) = \frac{R}{R+S}$ ist offensichtlich, da das Ziehen der ersten Kugel auf gut Glück erfolgt. \square

Index

- σ -Algebra, 15
- σ -Stetigkeit von Wahrscheinlichkeitsverteilungen, 27
- Algebra, 15
- Anfangsverteilung, 111
- Axiomensystem der Wahrscheinlichkeitstheorie, 24
- Bayes'sche Formel, 122
- bedingte Wahrscheinlichkeit, 118
 - im Laplace-Modell, 123
 - im mehrstufigen Versuch, 124
- Binomialverteilung, 77
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- Bonferroni-Ungleichungen, 30
- Borel-Cantelli
 - 1. Lemma von, 29
 - 2. Lemma von, 130
- Ein- und Ausschlussformel, 29
- Einpunktverteilung, 76
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- Ereignis
 - fast sicheres, 49
 - fast unmögliches, 49
 - zufälliges, 11
- Erwartungswert
 - diskret, 84, 86
- erzeugende Funktion, 102
- Exponentialverteilung
 - Verteilungsfunktion, 65
- geometrische Verteilung, 77
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- gleichmäßige Verteilung
 - diskret, 33
- gleichmäßige Verteilung, diskret, 76
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- hypergeometrische Verteilung, 81–83
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- Korrelationskoeffizient, 98
- Kovarianz, 98
- Laplace
 - Experiment, 32
- Münzenwurf, 16, 34
- Median, 58
- Moment
 - diskret, 87, 88

- diskret, zentriert, 87, 88
- Multiplikationssatz für Wahrscheinlichkeiten, 120
- negative Binomialverteilung, 78
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- Pfadregel
 - erste, 110
 - zweite, 110
- Poissonverteilung, 77
 - Erwartungswert, 85
 - erzeugende Funktion, 106
 - Varianz, 90
- Polya'sches Urnenschema, 111
- Quantil, 58
 - unteres, oberes, 58
- Randverteilung
 - diskret, 93
- Regressionsgerade, 101
- Streuung, *siehe* Varianz
- totale Wahrscheinlichkeit, Satz von, 121
- Uebergangsverteilung, 111
- Unabhängigkeit
 - in mehrstufigen Experimenten, 133
 - von σ -Algebren, 131
 - von Ereignissen, 127, 128
 - von Ereignissen, paarweise, 128
 - von Mengensystemen, 131
 - von Zufallsgrößen, 134, 136
- Ungleichung
 - von Cauchy-Schwarz, 97
 - von Tschebychev
 - diskret, 90
 - unkorreliert, 99
 - Urnenmodelle, 43–45
- Varianz
 - diskret, 89–90
- Verteilung
 - diskrete, 75
 - gemeinsame von U und V , diskret, 92
 - Wahrscheinlichkeits-, P^X , 52
- Verteilungsdichte, 63
- Verteilungsfunktion
 - der Zufallsgröße X , 55
 - diskret, 79
- Wahrscheinlichkeitsmaß, 24
- Wahrscheinlichkeitsraum, 25
- zufälliger Vektor, 53
 - diskret, zweidimensional, 91
 - Funktionen diskreter, 94
- zufälliger Versuch, 9
 - mehrstufig, 106–115
- Zufallsgröße, 51
 - diskret, 78
 - reellwertige, 53