

# Zusammenfassung Algorithmische Mathematik II

12. Juli 2013

## 0.1 Zufallsvariablen und ihre Verteilung

**Definition.** • Eine **diskrete Zufallsvariable** ist eine Abbildung

$$X : \Sigma \rightarrow S,$$

wobei  $S$  abzählbar sei.

- Die **Verteilung** von  $X$  ist die Wahrscheinlichkeitsverteilung  $\mu_X$  auf  $S$  mit Gewichten

$$p_X(a) := P[X^{-1}(a)]$$

**Bemerkung.** Wir schreiben  $\{X = a\}$  für  $X^{-1}(a)$  und  $P[X = a]$  statt  $P[\{X = a\}]$ .

Ist  $A \subseteq S$ , so kann  $\mu_X(A)$  als die Wahrscheinlichkeit interpretiert werden, mit der ein Element aus  $A$  ausgespuckt wird.

## 0.2 Binomialverteilung

Motivation: Man zieht eine Kugel aus einer Urne mit  $m$  Kugeln und legt sie wieder zurück. Das macht man  $n$  mal. Das mathematische Modell sieht folgendermaßen aus:

Die Kugeln werden mit  $1, 2, \dots, m$  durchnummeriert. Die Menge dieser Kugeln sei  $S := \{1, 2, \dots, m\}$

Der Ereignisraum ist dann  $\Omega = S^n$ , ein Elementarereignis ist dann  $(x_1, x_2, \dots, x_n) = \omega \in \Omega$ , wobei  $x_i \in S \forall i \in \{1, 2, \dots, n\}$ . (das sind die einzelnen Kugeln)

Es wird angenommen, dass die  $\omega$  gleichverteilt sind, dh. für alle  $\omega, \omega' \in \Omega$  gilt  $p(\omega) = p(\omega') = \frac{1}{|S|^n}$ .

Die Funktion  $X_i : \Omega \rightarrow S : \omega = (x_1, x_2, \dots, x_n) \rightarrow x_i$  gibt das  $i$ -te Ereignis zurück, also die Kugel, die als  $i$ -tes gezogen wurde.

Sei  $E \subseteq S$  ein Teil der  $m$  Kugeln mit einer besonderen Eigenschaft (schwarze Kugeln, etc.).

Die Wahrscheinlichkeit, dass beim  $i$ -ten Zug eine solche Kugel gezogen wird, ist gerade

$$P[x_i \in E] = \mu_{X_i}(E) = \frac{|E|}{|S|} =: p,$$

was als Erfolgswahrscheinlichkeit bezeichnet wird.

Die Wahrscheinlichkeit, dass dieser Erfolg  $k$  mal eintritt, wobei  $k \in \{1, \dots, n\}$ , ist

$$P[N = k] = \binom{n}{k} p^k (1-p)^{n-k} =: p_{n,p}(k)$$

Ist dies die Massenfunktion einer Wahrscheinlichkeitsverteilung auf  $\{0, \dots, n\}$ , so heißt diese Verteilung **Binomialverteilung** mit Parameter  $n$  und  $p$ . Sie gibt aus, mit welcher Wahrscheinlichkeit bei  $n$ -maligem Ziehen aus einer Urne genau  $k$  mal ein Erfolg gezogen wird. Für kleine Erfolgswahrscheinlichkeiten  $\frac{\lambda}{n}$  und große  $n$  nähert sich die Binomialverteilung an die **Poisson-**

**verteilung** mit Parameter  $\lambda$  an:

$$p(k) := \frac{\lambda^k}{k!} e^{-\lambda} = \lim_{n \rightarrow \infty} p_{n, \frac{\lambda}{n}}(k)$$

### 0.3 Hypergeometrische Verteilung

Motivation: Man zieht eine Kugel aus einer Urne mit  $m$  Kugeln ( $r$  rote,  $m - r$  schwarze) und legt sie nicht wieder zurück. Das macht man  $n$  mal. Das mathematische Modell ist im Wesentlichen analog zur Binomialverteilung. Für die Ereignisse gilt diesmal zusätzlich  $x_i \neq x_j \forall i, j \in \{1, \dots, m\}$ .  $N(\omega) :=$  Anzahl der roten Kugeln in  $\omega$ . Die Wahrscheinlichkeit, dass  $N(\omega) = k$  ist ( $k$  rote Kugeln in  $\omega$ ), ist

$$P[N = k] = \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}}$$

für  $k \in \{0, \dots, n\}$ . Diese Verteilung heißt **hypergeometrische Verteilung** mit Parametern  $m, r, n$ .

Für  $n \rightarrow \infty$  nähert sie sich an die Binomialverteilung an:

$$P[N = k] \rightarrow \binom{n}{k} p^k (1-p)^{n-k}$$

## 0.4 Bedingte Wahrscheinlichkeiten

**Definition.** Für Ereignisse  $A$  und  $B$  eines Wahrscheinlichkeitsraums  $(\Omega, \mathcal{A}, P)$  mit  $P[B] \neq 0$  heißt

$$P[A|B] := \frac{P[A \cap B]}{P[B]}$$

die **bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$**  („die Wahrscheinlichkeit dafür, dass  $A$  eintritt, wenn wir schon wissen, dass  $B$  eintritt“).

**Bemerkung.**

- $P[\bullet|B] : \mathcal{A} \rightarrow [0,1]$  ist eine Wahrscheinlichkeitsverteilung auf  $(\Omega, \mathcal{A})$ , die **bedingte Verteilung gegeben  $B$** .
- Der Erwartungswert  $E[X|B] = \sum_{a \in S} a \cdot P[X = a|B]$  einer diskreten Zufallsvariable  $X : \Omega \rightarrow S$  bezüglich der bedingten Verteilung heißt **bedingte Erwartung von  $X$  gegeben  $B$** .
- Im Fall der Gleichverteilung auf einer endlichen Menge gilt  $P[A|B] = \frac{|A \cap B|}{|B|}$ .

### 0.4.1 Berechnung von Wahrscheinlichkeiten durch Fallunterscheidung

Im Folgenden sei  $\Omega = \bigcup H_i$  eine disjunkte Zerlegung von  $\Omega$  in abzählbar viele Fälle („Hypothesen“).

**Satz.** Für alle  $A \in \mathcal{A}$  gilt  $P[A] = \sum_{i \in I, P[H_i] \neq 0} P[A|H_i] \cdot P[H_i]$ .

*Beweis.* Man verwendet die  $\sigma$ -Additivität und rechnet rum. □

Die Zerlegung in Hypothesen kann eventuell mehr Information als der Gesamtüberblick der Situation liefern (vgl. „Simpson-Paradoxon“ bei Bewerbungen in Berkeley).

### 0.4.2 Bayessche Regel

Wenn man wissen will, wie wahrscheinlich die Hypothesen  $H_i$  sind, kann man zuerst  $P[H_i]$  einschätzen („a priori degree of belief“). Wenn man dann zusätzlich weiß, dass ein Ereignis  $A \in \mathcal{A}$  mit  $P[A] \neq 0$  eintritt und die bedingte Wahrscheinlichkeit  $P[A|H_i]$  („likelihood“) für jedes  $H_i$  kennt, dann kann man eine neue Einschätzung („a posteriori degree of belief“) erhalten, und zwar gemäß dem folgenden

**Korollar.** (Bayessche Regel). Für  $A \in \mathcal{A}$  mit  $P[A] \neq 0$  gilt

$$P[H_i|A] = \frac{P[A|H_i] \cdot P[H_i]}{\sum_{j \in I, P[H_j] \neq 0} P[A|H_j] \cdot P[H_j]}$$

für alle  $i \in I$  mit  $P[H_i] \neq 0$ , d.h.  $P[H_i|A] = c \cdot P[H_i] \cdot P[A|H_i]$ , wobei  $c$  eine von  $i$  unabhängige Konstante ist.

(Man schätzt  $P[H_i]$  ein. Mit Hilfe von der obigen Formel rechnet man dann aus  $P[A]$  und den „likelihoods“ die „neue“ Einschätzung  $P[H_i|A]$ . Also erhält man theoretisch keine neue Information, aber diese „Wahrscheinlichkeiten“ sind häufig nur empirische Werte und in dem Fall kann man mit dieser Formel die bedingten Wahrscheinlichkeiten  $P[H_i|A]$  einschätzen.)

## 0.5 Mehrstufige diskrete Modelle

Für ein  $n$ -stufiges Zufallsexperiment mit abzählbaren Stichprobenräume  $\Omega_1, \dots, \Omega_n$  von Teilerperimenten kann  $\Omega = \prod \Omega_i$  als der Stichprobenraum des Gesamtexperiments aufgefasst werden.

Für  $\omega \in \Omega$  und  $k \in \{1, \dots, n\}$  sei  $X_k(\omega) = \omega_k$  der Ausgang des  $k$ -ten Experiments. Angenommen, wir kennen  $P[X_1 = x_1] = p_1(x_1)$  für alle  $x_1 \in \Omega_1$  und  $P[X_k = x_k | X_1 = x_1, \dots, X_{k-1} = x_{k-1}] = p_k(x_k | x_1, \dots, x_{k-1})$  für alle  $k \in \{1, \dots, n\}$ . Dann können wir die gesamte Wahrscheinlichkeitsverteilung  $P$  auf  $\Omega$  folgendermaßen erhalten:

**Satz.** Seien  $p_1$  und  $p_k(\bullet | x_1, \dots, x_{k-1})$  für jedes  $k = 2, \dots, n$  und  $x_i \in \Omega_i$  die Massenfunktion einer Wahrscheinlichkeitsverteilung auf  $\Omega_k$ . Dann existiert genau eine Wahrscheinlichkeitsverteilung  $P$  auf  $(\Omega, \mathcal{P}(\Omega))$ , die die obige Eigenschaften hat. Diese ist bestimmt durch die Massenfunktion

$$p(x_1, \dots, x_n) = p_1(x_1)p_2(x_2|x_1)p_3(x_3|x_1, x_2) \cdots p_n(x_n|x_1, \dots, x_{n-1}).$$

*Beweis.* Rumrechnerei. Die Eindeutigkeit folgt aus der Existenz.  $\square$

### 0.5.1 Produktmodelle

Ist der Ausgang des  $i$ -ten Experiments unabhängig von  $x_1, \dots, x_{i-1}$ , so gilt  $p_i(x_i | x_1, \dots, x_{i-1}) = p_i(x_i)$  mit einer von  $x_1, \dots, x_{i-1}$  unabhängigen Massenfunktion  $p_i$  einer Wahrscheinlichkeitsverteilung  $P_i$  auf  $\Omega_i$ . In diesem Fall gilt  $p(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i)$  für alle  $(x_1, \dots, x_n) \in \Omega$ .

**Definition.** Die Wahrscheinlichkeitsverteilung  $P$  auf  $\Omega = \prod \Omega_i$  mit obiger Massenfunktion heißt **Produkt** von  $P_1, \dots, P_n$  und wird mit  $P_1 \otimes \dots \otimes P_n$  notiert.

**Beispiel.** Bernouilliverteilung.

**Satz.** Im Produktmodell gilt für beliebige Ereignisse  $A_i \subseteq \Omega_i$

$$P[A_1 \times \dots \times A_n] = P[X_1 \in A_1, \dots, X_n \in A_n] = \prod_{i=1}^n P[X_i \in A_i] = \prod_{i=1}^n P_i[A_i],$$

d.h.  $X_1, \dots, X_n$  sind unabhängige Zufallsvariablen (s. nächstes Kapitel).

*Beweis.* Rechnung.  $\square$

### 0.5.2 Markov-Ketten

Wir wollen eine zufällige Entwicklung mit abzählbarem Zustandsraum  $S$  modellieren. Dazu betrachten wir den Stichprobenraum  $\Omega = S^{n+1}$ . Häufig hängt die Weiterentwicklung des Systems nur vom gegenwärtigen Zustand ab, d.h. es gilt  $p_k(x_k | x_0, \dots, x_{k-1}) = p_k(x_{k-1}, x_k)$  („Bewegungsgesetz“), wobei für  $p_k : S \times S \rightarrow [0, 1]$

1.  $p_k(x, y) \geq 0$  für alle  $x, y \in S$  und

$$2. \sum_{y \in S} p_k(x, y) = 1$$

gelten. (Dies bedeutet, dass  $p_k(x, \blacktriangle)$  für jedes  $x \in S$  die Massenfunktion einer Wahrscheinlichkeitsverteilung auf  $S$  ist.)

**Definition.** Eine „Matrix“  $p_k(x, y)$  mit den obigen Eigenschaften heißt **stochastische Matrix** (oder **stochastischer Kern**) auf  $S$ .

(Im Mehrstufenmodell gilt in dieser Situation  $p(x_0, \dots, x_n) = p_0(x_0)p_1(x_1, x_2) \cdots p_n(x_{n-1}, x_n)$ .) Der Fall, in dem  $p_k(x, y) = p(x, y)$  unabhängig von  $k$  ist, nennt man **zeitlich homogen**.

**Beispiel.**

- Produktmodell
- Random Walk auf  $\mathbb{Z}^d$
- Urnenmodelle

### 0.5.3 Berechnung von Wahrscheinlichkeiten

**Satz.** (*Markov-Eigenschaft*) Für alle  $0 \leq k < l \leq n$  und  $x_0, \dots, x_l \in S$  mit  $P[X_0 = x_0, \dots, X_k = x_k] \neq 0$  gilt

$$P[X_l = x_l | X_0 = x_0, \dots, X_k = x_k] = P[X_l = x_l | X_k = x_k] = (p_{k+1}p_{k+2} \cdots p_l)(x_k, x_l),$$

wobei  $(pq)(x, y) := \sum_{z \in S} p(x, z)q(z, y)$  das Produkt der Matrizen  $p$  und  $q$  ist.

*Beweis.* Indexschlacht und Rechnungskampf. □

## 0.6 Unabhängigkeit von Ereignissen

**Definition.** Zwei Ereignisse heissen unabhängig, falls

$$P[A \cap B] = P[A] \cdot P[B]$$

gilt.

Eine beliebige (nicht notwendig endlich oder abzählbar!) Kollektion von Ereignissen  $A_i$  ( $i \in I$ ) heisst unabhängig, falls

$$P[A_{i_1} \cap \dots \cap A_{i_n}] = \prod_{k=1}^n P[A_{i_k}] \text{ für alle } n \in \mathbb{N} \text{ und alle paarweise verschiedenen } i_1, \dots, i_n \in I$$

gilt.

**Satz.** Sind die Ereignisse  $A_1, \dots, A_n \in \mathcal{A}$  unabhängig und  $B_j = A_j$  oder  $B_j = A_j^C$  für alle  $j \in \{1, \dots, n\}$ , so sind auch die Ereignisse  $B_1, \dots, B_n$  unabhängig.

Seien  $A_1, A_2, \dots$  unabhängige Ereignisse mit jeweils Wahrscheinlichkeit  $p$ . Wir definieren die Wartezeit auf das erste Eintreten eines Ereignisses durch

$$T(\omega) = \min\{n \in \mathbb{N} : \omega \in A_n\}$$

.

Es gilt  $P[T = n] = p \cdot (1 - p)^{n-1}$ .

**Definition.** Die Wahrscheinlichkeitsverteilung auf  $\mathbb{N}$  mit Massenfunktion

$$p(n) = p \cdot (1 - p)^{n-1}$$

heisst geometrische Verteilung zum Parameter  $p$ .

Die Wahrscheinlichkeit, dass unter  $n$  Ereignissen  $k$  eintreten ist gleich der Binomialverteilung. Sei  $S_n$  gleich der Anzahl der eingetretenen Ereignisse innerhalb der ersten  $n$  Ereignisse.

**Satz.** (Bernstein-Ungleichung)

$$\forall \epsilon > 0 \forall n \in \mathbb{N} P\left[\frac{S_n}{n} \geq p + \epsilon\right] \leq e^{-2\epsilon^2 n}$$

(analog für  $\geq p - \epsilon$ )



## 0.7 Unabhängige Zufallsvariablen und Random Walk

**Definition.** Seien  $X_i : \Omega \rightarrow S_i, i \in \{1, \dots, n\}$  diskrete Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $(\Omega, A, P)$ . Dann ist  $(X_1, \dots, X_n) : \Omega \rightarrow S_1 \times \dots \times S_n$  eine Zufallsvariable.

Die Verteilung des Zufallsvektors  $(X_1, \dots, X_n)$  heißt **gemeinsame Verteilung** der Zufallsvariablen  $X_1, \dots, X_n$ . Ihre Massenfunktion ist

$$p_{X_1, \dots, X_n}(a_1, \dots, a_n) = P[X_1 = a_1, \dots, X_n = a_n].$$

Die diskreten Zufallsvariablen  $(X_1, \dots, X_n)$  heißen **unabhängig**, falls gilt

$$P[X_1 = a_1, \dots, X_n = a_n] = \prod_{i=1}^n P[X_i = a_i] \forall a_i \in S_i, i \in \{1, \dots, n\}$$

Unendlich viele diskrete Zufallsvariablen  $X_i : \Omega \rightarrow S_i, i \in I$  heißen **unabhängig**, falls die Ereignisse  $\{X_i = a_i\}, i \in I$  für alle  $a_i \in S$  unabhängig sind.

**Satz.** Die folgenden Aussagen sind äquivalent:

- $X_1, \dots, X_n$  sind unabhängig.
- $p_{X_1, \dots, X_n}(a_1, \dots, a_n) = \prod_{i=1}^n p_{X_i}(a_i)$
- $\mu_{X_1, \dots, X_n} = \mu_{X_1} \times \dots \times \mu_{X_n}$
- Die Ereignisse  $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$  sind unabhängig für alle  $A_i \subseteq S_i, i \in \{1, \dots, n\}$ .
- Die Ereignisse  $\{X_1 = a_1\}, \dots, \{X_n = a_n\}$  sind unabhängig für alle  $a_i \in S_i, i \in \{1, \dots, n\}$

Dabei wird immer wieder dieselbe Aussage getroffen, für einzelne Werte der Zufallsvariablen, oder für Teilmengen von Werten der Zufallsvariablen.

### 0.7.1 Der Random Walk auf $\mathbb{Z}$

Wir laufen auf der Zahlengeraden mit ganzzahligen Einträgen, beginnend beim Startwert  $a$ , mit Wahrscheinlichkeit  $p$  um 1 vorwärts und mit Wahrscheinlichkeit  $1 - p$  um 1 rückwärts. Die mathematische Modellierung ist wie folgt: Der Ereignisraum  $\Omega$  ist die Menge aller Random Walks, also alle Folgen  $(S_i)_{i \in \mathbb{N}}$ , mit  $S_0 = a \in \mathbb{Z}$  und  $|S_j - S_{j+1}| = 1 \forall j \in \mathbb{N}$ .

Der  $i - 1$ -te Schritt wird durch die Zufallsvariable  $X_i : \Omega \rightarrow \{-1, +1\}$  angegeben. Es gilt  $P[X_i = +1] = p, P[X_i = -1] = 1 - p \forall i \in \mathbb{N} \setminus \{0\}, p \in (0, 1)$ . Dann gilt  $S_0 = a, S_{n+1} = S_n + X_{n+1}$ . Induktiv folgt  $S_n = a + \sum_{i=1}^n X_i$ .

Klar ist, dass man in einer geraden Anzahl von Schritten stets ein Element aus  $a + 2\mathbb{Z}$  erreicht und in einer ungeraden Anzahl von Schritten stets ein Element aus  $a + 1 + 2\mathbb{Z}$  erreicht. Es gilt

**Lemma.** Sei  $k \in \mathbb{Z}$ . Dann gilt

$$P[S_n = a + k] = \begin{cases} 0 & \text{falls } n + k \text{ ungerade oder } |k| > n, \\ \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} (1-p)^{\frac{n-k}{2}} & \text{sonst.} \end{cases}$$

### 0.7.2 Symmetrischer Random Walk

Wir betrachten nun den Fall  $p = \frac{1}{2}$ .

Sei  $\lambda \in \mathbb{Z}$  fest. Wir definieren die Zufallsvariable

$$T_\lambda : \Omega \rightarrow \mathbb{N} \cup \{\infty\} : \omega \mapsto T_\lambda(\omega) := \inf\{n \in \mathbb{N} \setminus \{0\} \mid S_n(\omega) = \lambda\}$$

$T_\lambda(\omega)$  gibt den Zeitpunkt aus, an dem  $\lambda$  zum ersten Mal in  $\omega$  erreicht wird. Wir wollen  $P[T_\lambda \leq n] = P[\bigcup_{i=1}^n \{S_i = \lambda\}]$  berechnen, die Wahrscheinlichkeit, dass  $\lambda$  innerhalb der ersten  $n$  Schritte erreicht wird.

Nach  $n$  Schritten abbrechende Random Walks können als Folgen mit  $n$  Folgenglieder interpretiert werden, wobei wieder  $S_0 = a \in \mathbb{Z}$  und  $|S_j - S_{j+1}| = 1 \forall j \in \{0, \dots, n-1\}$  gilt. Bei gegebenem Startwert  $a$  gibt es genau  $2^n$  verschiedene solche Random Walks. Jede solche Folge tritt dabei mit gleicher Wahrscheinlichkeit auf.

**Satz.** *Reflektionsprinzip:* Seien  $\lambda, b \in \mathbb{Z}$ . Es gelte entweder ( $a < \lambda$  und  $b \leq \lambda$ ) oder ( $a > \lambda$  und  $b \geq \lambda$ ) (dh.  $a$  und  $b$  liegen beide rechts oder beide links von  $\lambda$ ). Dann gilt:

$$P[T_\lambda \leq n, S_n = b] = P[S_n = b^*],$$

wobei  $b^* := 2\lambda - b$  die Spiegelung von  $b$  an  $\lambda$  ist. (Dann muss ja  $2\lambda = b + b^*$  gelten)

Der Satz besagt also, dass wenn man bereits  $\lambda$  erreicht hat, dann ist die Wahrscheinlichkeit, dass man nach einem beliebigen Schritt insgesamt  $k$  Schritte vorwärts gegangen ist, gleich der Wahrscheinlichkeit, dass man nach einem beliebigen Schritt insgesamt  $k$  Schritte rückwärts gegangen ist.

**Satz.** (Trefferzeitenverteilung) Wir erinnern, dass  $a$  der Startwert des Random Walks ist. Es gilt

•

$$P[T_\lambda \leq n] = \begin{cases} P[S_n \geq \lambda] + P[S_n > \lambda], & \text{falls } \lambda > a \\ P[S_n \leq \lambda] + P[S_n < \lambda], & \text{falls } \lambda < a \end{cases}$$

•

$$P[T_\lambda = n] = \begin{cases} \frac{1}{2}P[S_{n-1} = \lambda - 1] - \frac{1}{2}P[S_{n-1} = \lambda + 1], & \text{falls } \lambda > a \\ \frac{1}{2}P[S_{n-1} = \lambda + 1] - \frac{1}{2}P[S_{n-1} = \lambda - 1], & \text{falls } \lambda < a \end{cases} = \begin{cases} \frac{\lambda-a}{n} \binom{n}{\frac{n+\lambda-a}{2}} 2^{-n}, & \text{falls } \lambda > a \\ \frac{a-\lambda}{n} \binom{n}{\frac{n+\lambda-a}{2}} 2^{-n}, & \text{falls } \lambda < a \end{cases}$$

**Korollar.** (*Verteilung des Maximums*) Sei  $M_n := \max\{S_0, \dots, S_n\}$ . Für  $\lambda > a$  gilt

$$P[M_n \geq \lambda] = P[T_\lambda \leq n] = P[S_n \geq \lambda] + P[S_n > \lambda]$$

## 0.8 Varianz und Kovarianz

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow S$  eine Zufallsvariable auf  $(\Omega, \mathcal{A}, P)$ , so dass  $E[|X|]$  endlich ist.

**Definition.**

$$\text{Var}(X) := E[(X - E[X])^2]$$

heißt **Varianz** von  $X$  und liegt in  $[0, \infty]$ .

$$\sigma(X) := \text{Var}(X)^{1/2}$$

heißt **Standardabweichung** von  $X$ .

Interpretation: Kennzahl für die Größe der Fluktuationen von  $X$  um  $E[X]$ ; Maß für Risiko bei Prognose des Ausgangs  $X(\omega)$  durch  $E[X]$ .

**Bemerkung.** •  $\text{Var}(X) = \sum_{a \in S} (a - m)^2 p_X(a)$ , wobei  $m = E[X] = \sum_{a \in S} a \sum_a p_X(a)$ .

- $\text{Var}(X) = 0$  gdw.  $P[X = E[X]] = 1$ .
- $\text{Var}(X) = E[X^2] - E[X]^2$ .
- $\text{Var}(aX + b) = \text{Var}(aX) = a^2 \text{Var}(X)$ .

**Beispiel.** • Sei  $X = 1$  mit Wahrscheinlichkeit  $p$  und  $X = 0$  mit Wahrscheinlichkeit  $1 - p$ . Dann ist  $\text{Var}(X) = p(1 - p)$ .

- Sei  $T$  geometrisch verteilt mit Parameter  $p \in (0, 1]$ . Dann ist  $\text{Var}(T) = \frac{1-p}{p^2}$ .

**Definition.**

$$\mathcal{L}^2(\Omega, \mathcal{A}, P) := \{X : \Omega \rightarrow \mathbb{R} \mid E[X^2] < \infty\}$$

**Lemma.** • Für Zufallsvariablen  $X, Y \in \mathcal{L}^2$  gilt:  $E[|XY|] \leq E[X^2]^{1/2} E[Y^2]^{1/2} < \infty$

- $\mathcal{L}^2$  ist ein Vektorraum und  $(X, Y)_{\mathcal{L}^2} := E[XY]$  ist eine positiv semidefinite symmetrische Bilinearform (Skalarprodukt) auf  $\mathcal{L}^2$ . Insbesondere gilt die Cauchy-Schwarz-Ungleichung.
- Für  $X \in \mathcal{L}^2$  gilt  $E[|X|] < \infty$

**Definition.** Seien  $X, Y \in \mathcal{L}^2$ .

- $\text{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$  heißt **Kovarianz** von  $X$  und  $Y$ .
- Gilt  $\sigma(X), \sigma(Y) \neq 0$ , so heißt  $\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$  **Korrelationskoeffizient** von  $X$  und  $Y$ .
- $X$  und  $Y$  heißen **unkorreliert**, falls  $\text{Cov}(X, Y) = 0$ , d.h. falls  $E[XY] = E[X] \cdot E[Y]$ .

**Satz.** Seien  $X : \Omega \rightarrow S$  und  $Y : \Omega \rightarrow T$  diskrete Zufallsvariablen auf  $(\Omega, \mathcal{A}, P)$ . Dann sind äquivalent:

- $X$  und  $Y$  sind unabhängig
- $f(X)$  und  $g(Y)$  sind unkorreliert für alle Funktionen  $f : S \rightarrow \mathbb{R}$  und  $g : T \rightarrow \mathbb{R}$  mit  $f(X), g(Y) \in \mathcal{L}^2$ .

**Beispiel.** Sei  $X = 1, 0, -1$ , jeweils mit Wahrscheinlichkeit  $\frac{1}{3}$ , und  $Y = X^2$ . Dann sind  $X$  und  $Y$  nicht unabhängig, aber unkorreliert. Intuition: Unkorreliertheit bedeutet nur kein linearer Zusammenhang. Hier liegt ein quadratischer Zusammenhang vor.

**Satz.** Für  $X_1, \dots, X_n \in \mathcal{L}^2$  gilt:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i,j=1, i < j}^n \text{Cov}(X_i, X_j)$$

## 0.9 Monte Carlo-Verfahren

Sei  $S$  eine abzählbare Menge und  $\mu$  eine Wahrscheinlichkeitsverteilung auf  $S$ . Im folgenden bezeichnen wir auch die Massenfunktion mit  $\mu$ , d.h.  $\mu(x) := \mu(\{x\})$  für alle  $x \in S$ .

Sei  $f : S \rightarrow \mathbb{R}$  mit  $E_\mu[f^2] = \sum_{x \in S} f(x)^2 \mu(x) < \infty$ . Dann kann man den Erwartungswert  $\theta := E_\mu[f] = \sum_{x \in S} f(x) \mu(x)$  durch die *Monte Carlo-Schätzer*

$$\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^n f(X_i)$$

approximieren, wobei  $X_1, \dots, X_n$  unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit Verteilung  $\mu$  sind. Die Abschätzung aus dem Gesetz der großen Zahlen gibt uns für diese Folge das folgende

**Korollar.**  $P[|\theta - \hat{\theta}_n| \geq \varepsilon] \leq \frac{1}{n\varepsilon^2 \text{Var}_\mu[f]} \rightarrow 0$  für  $n \rightarrow \infty$ , d.h.  $\hat{\theta}_n$  ist eine **konsistente Schätzfolge** für  $\theta$ .

**Bemerkung.**

- $\hat{\theta}_n$  ist ein erwartungstreuer Schätzer:

$$E[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^n E[f(X_i)] = E_\mu[f] = \theta,$$

- Für den mittleren quadratischen Fehler gilt

$$E[|\theta - \hat{\theta}_n|^2] = \text{Var}(\hat{\theta}_n) = \frac{1}{n} \text{Var}_\mu[f],$$

also insbesondere  $\|\theta - \hat{\theta}_n\|_{\mathcal{L}^2} = \sqrt{E[|\theta - \hat{\theta}_n|^2]} = O(1/\sqrt{n})$ .

**Beispiel.** Sei  $B \subseteq S$ . Für  $p = \mu(B) = E_\mu[I_B]$  ist dann  $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n I_B(X_i)$  ein Monte Carlo-Schätzer.

**Bemerkung.** Für kleine Wahrscheinlichkeiten braucht dieses einfache Monte Carlo-Verfahren sehr viele Stichproben, wenn man die Wahrscheinlichkeit mit einem kleinem Fehler bestimmen will.

### 0.9.1 Varianzreduktion durch Importance Sampling

Sei  $\nu$  eine weitere Wahrscheinlichkeitsverteilung auf  $S$  mit  $\nu(x) > 0$  für alle  $x \in S$ . Dann kann man  $\theta$  auch bezüglich  $\nu$  ausdrücken:

$$\theta = E_\mu[f] = \sum_{x \in S} f(x) \mu(x) = \sum_{x \in S} f(x) \frac{\mu(x)}{\nu(x)} \nu(x) = E_\nu[f\rho],$$

wobei  $\rho(x) = \frac{\mu(x)}{\nu(x)}$ .

Ein alternativer Monte Carlo-Schätzer für  $\theta$  ist folglich  $\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n f(Y_i) \rho(Y_i)$ , wobei die  $Y_i$  unabhängige Zufallsvariablen mit Verteilung  $\nu$  sind.

$\tilde{\theta}_n$  ist ebenfalls erwartungstreu. Für die Varianz erhält man

$$\text{Var}(\tilde{\theta}_n) = \frac{1}{n} \text{Var}_\nu(f\rho) = \frac{1}{n} \left( \sum_{x \in S} f(x)^2 \rho(x)^2 \nu(x) - \theta^2 \right).$$

Bei geeigneter Wahl von  $\nu$  kann also die Varianz von  $\tilde{\theta}_n$  kleiner sein als die von  $\hat{\theta}_n$ . ( $\nu(x)$  soll groß sein, wenn  $|f(x)|$  groß ist.)