Anhang F

Zufallsvariablen

In diesem Anhang werden einige Elemente der Wahrscheinlichkeitstheorie dargestellt, insbesondere der Theorie der Zufallsvariablen, unter dem Gesichtspunkt eines an praktischen Fakten interessierten Physikers.

F.1 Definition

Der Begriff der *Zufallsvariable* (auch *stochastische Variable* oder *Zufallsgröße* genannt) X beruht auf zwei Elementen:

a) Die Menge (*Ergebnismenge*, *Zustandsmenge*) Ω der möglichen Werten (*Realisationen*) x der Ergebnisse eines Zufallsexperiments.

Diese Menge kann diskret oder kontinuierlich sein, oder teilweise diskret und teilweise kontinuierlich. Außerdem kann die Zustandsmenge mehrdimensional sein. Dementsprechend spricht man jeweils von einer diskreten, kontinuierlichen (oder stetigen), oder mehrdimensionalen Zufallsvariable. Die Letzteren werden hiernach als Vektoren $\boldsymbol{X}$ bezeichnet.

Ein Beispiel einer diskreten bzw. stetigen eindimensionalen Zufallsvariable ist die Projektion des Spins eines Teilchens auf eine gegebene Achse bzw. die kinetische Energie eines freien Teilchens. Die drei Komponenten der Geschwindigkeit eines Brown'schen[ai] Teilchens bilden eine kontinuierliche dreidimensionale Zufallsvariable. In der statistischen Feldtheorie sind die Zufallsgrößen Felder, die kontinuierliche unendlich dimensionale Zufallsvariablen bilden.

[ai] R. Brown, 1773–1858

b) Die Wahrscheinlichkeitsverteilung über dieser Menge.

Im Fall einer eindimensionalen kontinuierlichen Zufallsvariable mit Werten in einem Intervall (oder einer Vereinigung von Intervallen) $\mathcal{I}$ wird diese Verteilung durch eine *Wahrscheinlichkeitsdichte(funktion)* gegeben. Dabei handelt es sich um eine nicht-negative Funktion p_X

$$p_X(x) \geq 0 \quad \forall x \in \mathcal{I}, \tag{F.1a}$$

die auf 1 normiert ist:

$$\int_{\mathcal{I}} p_X(x)\,\mathrm{d}x = 1. \tag{F.1b}$$

$p_X(x)\,\mathrm{d}x$ ist die Wahrscheinlichkeit dafür, dass X einen Wert zwischen x und $x+\mathrm{d}x$ annimmt.

Dies lässt sich einfach auf den Fall einer mehrdimensionalen kontinuierlichen Zufallsvariable $\boldsymbol{X}$ mit Werten in einem Gebiet $\mathcal{G}$ verallgemeinern:

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0 \quad \forall \boldsymbol{x} \in \mathcal{G} \qquad \text{und} \qquad \int_{\mathcal{G}} p_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}^D\boldsymbol{x} = 1, \tag{F.2}$$

wobei das D-dimensionale Elementarvolumen um einen Punkt $\boldsymbol{x}$ als $\mathrm{d}^D\boldsymbol{x}$ bezeichnet wird.

Die Wahrscheinlichkeitsdichte kann auch Dirac-Distributionen beinhalten:

$$p_X(x) = \sum_n p_n \delta(x - x_n) + \tilde{p}_X(x), \tag{F.3a}$$

mit der Normierungsbedingung

$$\sum_n p_n + \int \tilde{p}_X(x)\,\mathrm{d}x = 1, \tag{F.3b}$$

mit $p_n > 0$ und einer nicht-negativen Funktion $\tilde{p}_X$. Im Fall $\tilde{p}_X = 0$ ist X einfach eine diskrete Zufallsvariable. Die zugehörige Wahrscheinlichkeitsdichte wird dann eher durch eine diskrete *Wahrscheinlichkeitsfunktion* ersetzt, die die endlichen Wahrscheinlichkeiten p_n für die einzelnen Werte x_n angibt.

In der formellen Wahrscheinlichkeitstheorie unterscheidet man zwischen der Ergebnismenge (d.h. die Menge aller *Elementarereignisse*) Ω und der Ereignisalgebra $\mathcal{F}$, die eine Teilmenge der Potenzmenge (Menge aller Teilmengen) von Ω darstellt. Die Elemente von $\mathcal{F}$ werden *Ereignisse* genannt, und beschreiben alle... Ereignisse, die man betrachten kann. Dann wird eine Funktion (*Wahrscheinlichkeitsmaß*) $\mathcal{P}$ von $\mathcal{F}$ in das reelle Intervall $[0, 1]$ eingeführt, die jedem Ereignis A eine Wahrscheinlichkeit $\mathcal{P}(A)$ zuordnet, mit den Bedingungen

- $\mathcal{P}(\Omega) = 1$ [Normierungseigenschaft, vgl. Gl. (F.1b) oder (F.3b)],
- $\forall A, B \in \mathcal{F}$, $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$ falls $\mathcal{P}(A \cap B) = 0$ — d.h. insbesondere wenn $A \cap B = \emptyset$ — sonst $\mathcal{P}(A \cup B) < \mathcal{P}(A) + \mathcal{P}(B)$.

Das Tripel $(\Omega, \mathcal{F}, \mathcal{P})$ wird *Wahrscheinlichkeitsraum* genannt.

Sei ein solcher Wahrscheinlichkeitsraum. Eine eindimensionale Zufallsvariable ist eine Funktion von Ω in $\mathbb{R}$ mit der Eigenschaft

$$\forall x \in \mathbb{R}, \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F},$$

d.h. die Menge aller Elementarereignisse ω, deren Realisation $X(\omega)$ kleiner als x ist, bildet ein Ereignis. Dieser Zufallsvariable wird die *Verteilungsfunktion* F von $\mathbb{R}$ in $[0, 1]$ assoziiert, die der reellen Zahl x die Wahrscheinlichkeit $\mathcal{P}(X \leq x) \equiv \mathcal{P}(\{\omega \in \Omega \mid X(\omega) \leq x\})$ zuordnet. Dann gilt

$$F(x) \equiv \mathcal{P}(X \leq x) = \int_{-\infty}^{x^+} p_X(x')\,\mathrm{d}x',$$

mit der Wahrscheinlichkeitsdichte $p_X(x)$. (Hier bedeutet die Bezeichnung x^+, dass falls p_X einen $\delta(x)$-Beitrag enthält, dann wird dieser im Integral mitgezählt.)

Bemerkung: Physikalische Größen haben oft eine Dimension: Länge, Masse, Zeit... Somit soll die Wahrscheinlichkeitsdichte p_G für die Verteilung der Werte g einer solchen Größe G ebenfalls eine

Dimension haben, und zwar die umgekehrte Dimension von G, damit die Wahrscheinlichkeit $p_G(g)\,\mathrm{d}g$ dimensionslos sei. Dies lässt sich auf die Wahrscheinlichkeitsdichten in Abschn. F.3 prüfen.

F.2 Erwartungswerte und Momente

Neben der Ergebnismenge Ω und der Wahrscheinlichkeitsdichte p_X werden weitere Begriffe eingeführt, um eine Zufallsvariable X zu charakterisieren. In diesem Abschnitt ist X eindimensional.

Sei f eine Funktion definiert auf Ω. Ihr *Erwartungswert* wird gegeben durch

$$\boxed{\langle f(X)\rangle \equiv \int_\Omega f(x)\, p_X(x)\,\mathrm{d}x.} \tag{F.4}$$

Bemerkungen:

∗ Dieser Erwartungswert wird auch (insbesondere in der mathematischen Literatur) als $E\big(f(X)\big)$ oder $E\big[f(X)\big]$ geschrieben.

∗ Das Bilden des Erwartungswerts ist eine lineare Operation.

Sei $m \in \mathbb{N}$. Das *m-te Moment* (oder *Moment der Ordnung m*) einer eindimensionalen Zufallsvariable X (bzw. der zugehörigen Wahrscheinlichkeitsverteilung) ist der Erwartungswert

$$\mu_m \equiv \langle X^m\rangle\,. \tag{F.5}$$

Insbesondere ist μ_1 der Erwartungswert der Zufallsvariable, der in Analogie zum arithmetischem Mittel oft „Mittelwert" genannt wird.

Außerdem wird die *Varianz* der Wahrscheinlichkeitsverteilung definiert durch

$$\sigma^2 \equiv \big\langle (X-\langle X\rangle)^2\big\rangle = \mu_2 - \mu_1^2. \tag{F.6}$$

Die positive Quadratwurzel σ heißt *Standardabweichung*. Die Letztere wird oft (aber ungenau) „Fluktuation" genannt, weil σ ein Maß für die Streuung der Zufallsvariable um ihren Erwartungswert bildet, d.h. für die Skala der Fluktuationen der durch die Zufallsvariable beschriebene Größe.

Bemerkungen:

∗ Das Integral, das das m-te Moment einer Wahrscheinlichkeitsverteilung ergibt, kann divergieren! Vgl. z.B. § F.3.8.

∗ Ähnlich der Varianz kann man auch das m-te *zentrale Moment* $\langle (X-\langle X\rangle)^m\rangle$ für eine beliebige $m \in \mathbb{N}$ definieren.

∗ Wenn die Zufallsvariable eine physikalische Dimension hat, gilt das auch für ihre Momente.

Ein weiterer nützlicher Begriff ist der der *charakteristischen Funktion*, definiert für $k \in \mathbb{R}$ durch

$$G_X(k) \equiv \left\langle \mathrm{e}^{\mathrm{i}kX}\right\rangle = \int_\Omega \mathrm{e}^{\mathrm{i}kx} p_X(x)\,\mathrm{d}x. \tag{F.7a}$$

Man prüft einfach, dass die Taylor-Entwicklung von $G(k)$ im Ursprungspunkt $k = 0$ lautet

$$G_X(k) = \sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!}\mu_m, \tag{F.7b}$$

d.h. die m-te Ableitung der charakteristischen Funktion im Punkt $k = 0$ ist mit dem m-ten Moment der Wahrscheinlichkeitsverteilung verknüpft.

Bemerkungen:

∗ Genauer werden die Momente durch die sukzessiven Ableitungen in $k = 0$ der *momenterzeugenden Funktion* $\mathcal{G}_X(k) \equiv G_X(-\mathrm{i}k)$ gegeben.

∗ Dank der Normierungseigenschaft (F.1b) ist die charakteristische Funktion $G(k)$ für beliebigen reellen k definiert — das Integral ist immer konvergent. Dagegen ist es nicht immer der Fall der momenterzeugenden Funktion.

∗ Die Analytizität der charakteristischen Funktion in $k=0$ hängt von der Existenz der Momente μ_m ab, d.h. Gl. (F.7b) gilt nur dann, wenn alle Momente beliebiger Ordnung existieren.

∗ Der Logarithmus von G (oder $\mathcal{G}$) erzeugt die sukzessiven *Kumulanten* κ_m der Verteilung:

$$\ln G_X(k) = \sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!}\kappa_m, \tag{F.8}$$

die manchmal nützlicher als die Momente sind (s. Abschn. F.4). Es gelten z.B. $\kappa_1 = \mu_1 = \langle X\rangle$ und $\kappa_2 = \sigma^2$.

F.3 Spezielle Wahrscheinlichkeitsverteilungen

In diesem Abschnitt werden einige Wahrscheinlichkeitsverteilungen dargestellt, die in der Physik oft auftauchen, anfangend mit diskreten Verteilungen, gefolgt durch kontinuierliche Dichten.

F.3.1 Diskrete Gleichverteilung

Sei X eine Zufallsvariable mit der endlichen diskreten Ergebnismenge $\Omega = \{x_1,\dots,x_N\}$ mit $N\in\mathbb{N}$. Die *diskrete Gleichverteilung*

$$p_n = \frac{1}{N} \quad \forall n \in \{1,2,\dots,N\} \tag{F.9}$$

beschreibt den Fall, in dem die Wahrscheinlichkeit für alle Realisationen gleich ist.

Der Erwartungswert ist $\langle X\rangle = \frac{1}{N}\sum_{n=1}^{N} x_n$ und die Varianz $\sigma^2 = \frac{1}{N^2}\sum_{n=1}^{N} x_n^2 - \frac{1}{N^2}\left(\sum_{n=1}^{N} x_n\right)^2$.

F.3.2 Bernoulli-Verteilung

Zufallsexperimente, bei denen es nur zwei mögliche Ergebnisse gibt — oder genauer, deren möglichen Ergebnisse in nur zwei Gruppen aufgeteilt werden —, werden *Bernoulli*[(aj)]*-Experimente* genannt. Die zwei möglichen Ergebnisse werden oft mit „Erfolg“ und „Misserfolg“ bezeichnet.

Eine Wahrscheinlichkeitsverteilung mit $\Omega = \{\text{Erfolg}, \text{Misserfolg}\}$ und den jeweiligen Wahrscheinlichkeiten p und $1-p$, wobei $0<p<1$, heißt *Bernoulli-Verteilung* mit dem Parameter p.

Mit der (üblichen) Zuordnung Erfolg $=1$ und Misserfolg $=1$ ist der Erwartungswert $\langle X\rangle = p$ und die Varianz $\sigma^2 = p(1-p)$.

F.3.3 Binomialverteilung

Sei p eine reelle Zahl, $0<p<1$, und $N\in\mathbb{N}$.

Die *Binomial-Verteilung* zu den Parametern N und p ist die Wahrscheinlichkeitsverteilung für eine Zufallsvariable mit der Ergebnismenge $\Omega = \{1,2,\dots,n,\dots N\}$ gegeben durch

$$p_n = \binom{N}{n} p^n (1-p)^{N-n}. \tag{F.10}$$

p_n ist die Wahrscheinlichkeit, dass es in der N-fachen Durchführung eines Bernoulli-Experiments genau n „Erfolge“ gibt.

Der Erwartungswert ist $\langle X\rangle = pN$ und die Varianz $\sigma^2 = Np(1-p)$.

(aj) J. Bernoulli, 1655–1705

Bemerkung: Eine verwandte Verteilung ist die sog. *negative Binomialverteilung* oder *Pascal*[ak]*-Verteilung* mit zwei Parameter $r \in \mathbb{N}$ — man kann die Definition auf einen postiven reellen r verallgemeinern — und $0 < p < 1$, mit der Ergebnismenge $\Omega = \{0, 1, 2, \ldots, \}$ und den zugehörigen Wahrscheinlichkeiten

$$p_n = \binom{n+r-1}{n} p^r (1-p)^n. \tag{F.11}$$

p_n ist die Wahrscheinlichkeit, dass es in einem $(n+r)$-fach wiederholten Bernoulli-Experiment nach n Misserfolgen genau r Erfolge gibt. Der Erwartungswert ist $\langle X \rangle = r(1-p)/p$ und die Varianz $\sigma^2 = r(1-p)/p^2$.

F.3.4 Poisson-Verteilung

Sei λ eine positive reelle Zahl. Die *Poisson*[al]*-Verteilung* zum Parameter λ ordnet dem natürlichen Zahl $n \in \mathbb{N} = \Omega$ (Ergebnismenge) die Wahrscheinlichkeit

$$p_n = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \tag{F.12}$$

zu. Der Erwartungswert und die Varianz sind $\langle X \rangle = \sigma^2 = \lambda$.

F.3.5 Stetige Gleichverteilung

Sei X eine kontinuierliche Zufallsvariable, deren Ergebnismenge Ω das reelle Intervall $]a, b]$ mit $a < b$ ist. Eine konstante Wahrscheinlichkeitsdichte

$$p_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{für } a < x \leq b \\ 0 & \text{sonst} \end{cases} \tag{F.13}$$

auf dem Intervall definiert die *stetige Gleichverteilung*, auch Uniformverteilung genannt. Dies stellt offenbar die Verallgemeinerung der diskreten Gleichverteilung zum Fall einer kontinuierlichen Ergebnismenge dar.

F.3.6 Gauß-Verteilung

Sei X eine kontinuierliche Zufallsvariable mit der Ergebnismenge $\Omega = \mathbb{R}$, und zwei reelle Zahlen μ und $\sigma > 0$. Die Wahrscheinlichkeitsdichte

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{F.14}$$

heißt *Gauß-Verteilung* (oder „Normalverteilung“).

Der Erwartungswert ist $\langle X \rangle = \mu$ und die Varianz σ^2. Außerdem kann man einfach prüfen, dass die Kumulanten κ_m aller Ordnungen $m \geq 3$ null sind.

F.3.7 Exponentialverteilung

Sei λ eine positive reelle Zahl. Eine kontinuierliche Zufallsvariable X mit der Ergebnismenge $\Omega = \mathbb{R}_+$ genügt der *Exponentialverteilung* mit Parameter λ, wenn ihre Wahrscheinlichkeitsdichte durch

$$p_X(x) = \lambda\, \mathrm{e}^{-\lambda x} \tag{F.15}$$

gegeben ist. Der Erwartungswert ist $\langle X \rangle = \dfrac{1}{\lambda}$ und die Varianz $\sigma^2 = \dfrac{1}{\lambda^2}$.

[ak] B. Pascal, 1623–1662 [al] S. D. Poisson, 1781–1840

F.3.8 Cauchy–Lorentz-Verteilung

Seien x_0 und γ zwei reelle Zahlen, mit $\gamma > 0$. Die *Cauchy*[am]*–Lorentz*[an]*-Verteilung*, die in der Physik auch oft *Breit*[ao]*–Wigner*[ap]*-Verteilung* genannt wird, ist

$$p_X(x) = \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \tag{F.16}$$

für $x \in \Omega = \mathbb{R}$.

Alle Momente dieser Verteilung divergieren! x_0 ist die Position des Maximums der Verteilung, während 2γ die volle Breite der Kurve auf halber Maximalhöhe darstellt.

F.4 Mehrdimensionale Wahrscheinlichkeitsverteilungen

Sei $\boldsymbol{X}$ eine D-dimensionale Zufallsgrößen, deren Komponenten als $X_1, X_2, \ldots, X_D$ bezeichnet werden. Der Kürze halber wird hiernach nur der Fall einer stetigen mehrdimensionalen Zufallsvariablen betrachtet.

F.4.1 Definitionen

Die Wahrscheinlichkeitsdichte $p_D(x_1, \ldots, x_D) \equiv p_{\boldsymbol{X}}(\boldsymbol{x})$ wird auch multivariate oder *gemeinsame* Wahrscheinlichkeitsdichte der D Zufallsvariablen genannt.

Sei $d < D$ und d Zufallsvariablen unten $X_1, X_2, \ldots, X_D$ — der Einfachheit halber, die d ersten $X_1, \ldots, X_d$. Die Wahrscheinlichkeit, dass sie Werte in den jeweiligen Intervallen $[x_1, x_1 + \mathrm{d}x_1]$, …, $[x_d, x_d + \mathrm{d}x_d]$ annehmen, *unabhänhig* von den Werten von $X_{d+1}, \ldots, X_D$, ist

$$p_d(x_1, \ldots, x_d)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d = \left[\int p_D(x_1, \ldots, x_d, x_{d+1}, \ldots, x_D)\, \mathrm{d}x_{d+1} \cdots \mathrm{d}x_D \right] \mathrm{d}x_1 \cdots \mathrm{d}x_d,$$

wobei das Integral über die Ergebnismengen für $X_{d+1}, \ldots, X_D$ läuft. Dann wird

$$p_d(x_1, \ldots, x_d) \equiv \int p_D(x_1, \ldots, x_d, x_{d+1}, \ldots, x_D)\, \mathrm{d}x_{d+1} \cdots \mathrm{d}x_D \tag{F.17}$$

Marginalverteilung oder *Randverteilung* genannt.

Wenn man den Zufallsvariablen $X_{d+1}, \ldots, X_D$ feste Werte $x_{d+1}, \ldots, x_D$ zuordnet, kann man die (bedingte) Wahrscheinlichkeitsverteilung für die anderen Zufallsvariablen unter dieser Bedingung betrachten. Somit führt man die *bedingte Wahrscheinlichkeitsdichte*

$$p_{d|D-d}(x_1, \ldots, x_d \,|\, x_{d+1}, \ldots, x_D) \tag{F.18}$$

ein. Es gilt dann

$$\begin{aligned} p_D(x_1, \ldots, x_D) &= p_{D-d}(x_{d+1}, \ldots, x_D)\, p_{d|D-d}(x_1, \ldots, x_d \,|\, x_{d+1}, \ldots, x_D) \\ &= p_d(x_1, \ldots, x_d)\, p_{D-d|d}(x_{d+1}, \ldots, x_D \,|\, x_1, \ldots, x_d). \end{aligned}$$

Die erste Gleichheit kann auch als

$$\boxed{p_{d|D-d}(x_1, \ldots, x_d \,|\, x_{d+1}, \ldots, x_D) = \frac{p_D(x_1, \ldots, x_D)}{p_{D-d}(x_{d+1}, \ldots, x_D)}} \tag{F.19}$$

umgeschrieben werden, entsprechend dem *Satz von Bayes*[aq].

Die *Momente* einer mehrdimensionalen Wahrscheinlichkeitsverteilung werden definiert durch

$$\langle X_1^{m_1} X_2^{m_2} \cdots X_D^{m_D} \rangle \equiv \int x_1^{m_1} x_2^{m_2} \cdots x_D^{m_D}\, p_D(x_1, \ldots, x_D)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_D. \tag{F.20}$$

[am] A. L. Cauchy, 1789–1857 [an] H. A. Lorentz, 1853–1926 [ao] G. Breit, 1899–1981 [ap] E. Wigner, 1902–1995
[aq] T. Bayes, 1701–1761

Sie werden erzeugt durch die *charakteristische Funktion*

$$G_{\boldsymbol{X}}(k_1,\dots,k_D) \equiv \left\langle \mathrm{e}^{\mathrm{i}(k_1 X_1+\cdots+k_D X_D}\right\rangle, \tag{F.21}$$

mit reellen Hilfsvariablen $k_1,\dots,k_D$. Dann erzeugt der Logarithmus der charakteristischen Funktion die Kumulanten. Insbesondere kommt zur zweiten Ordnung — als Koeffizient (bis auf einen Faktor i^2) von $k_j k_l$ mit $l \neq j$ — die *Kovarianz* der Zufallsvariablen X_j und X_l, definiert durch

$$\mathrm{Cov}(X_j, X_l) \equiv \big\langle \big(X_j - \langle X_j\rangle\big)\big(X_l - \langle X_l\rangle\big)\big\rangle\,. \tag{F.22}$$

F.4.2 Stochastische Unabhängigkeit

Wenn für alle Realisationen $x_1,\dots,x_d,x_{d+1},\dots,x_D$ der Zufallsvariablen die Gleichheit

$$p_D(x_1,\dots,x_D) = p_d(x_1,\dots,x_d)\, p_{D-d}(x_{d+1},\dots,x_D) \tag{F.23}$$

gilt, werden die Variablensätze $\{X_1,\dots,X_d\}$ und $\{X_{d+1},\dots,X_D\}$ als *stochastisch unabhängig* (oft kurz nur „unabhängig“) bezeichnet. Dann ist die Marginalverteilung (F.17) gleich der bedingten Verteilung (F.18).

Es seien z.B. zwei unabhängige Zufallsvariablen X_1 und X_2. Dann gilt für alle Funktionen f_1, f_2 definiert auf den jeweiligen Ergebnismengen[18] $\langle f_1(X_1)\, f_2(X_2)\rangle = \langle f_1(X_1)\rangle\, \langle f_2(X_2)\rangle$. Insbesondere gilt für alle Momente

$$\langle X_1^{m_1} X_2^{m_2}\rangle = \langle X_1^{m_1}\rangle\, \langle X_2^{m_2}\rangle \quad \forall m_1, m_2. \tag{F.24}$$

Bemerkung: Wenn die Kovarianz (F.22) zweier Zufallsvariablen null ist, werden sie *unkorreliert* genannt. Unabhängige Zufallsvariablen sind laut Gl. (F.24) mit $m_1 = m_2 = 1$ unkorreliert, die Umkehrung gilt aber nicht.

F.4.3 Summe von Zufallsvariablen

Es seien wieder zwei Zufallsvariablen X_1 und X_2, definiert auf derselben Ergebnismenge, mit der gemeinsamen Wahrscheinlichkeitsdichte $p_{\boldsymbol{X}}(x_1, x_2)$.

Ihre Summe $Y = X_1 + X_2$ bildet eine neue Zufallsvariable, mit dem Erwartungswert

$$\langle Y\rangle = \langle X_1\rangle + \langle X_2\rangle \tag{F.25}$$

und der Wahrscheinlichkeitsdichte

$$\begin{aligned} p_Y(y) &= \int p_{\boldsymbol{X}}(x_1,x_2)\, \delta(y - x_1 - x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= \int p_{\boldsymbol{X}}(x_1, y - x_1)\, \mathrm{d}x_1 \int p_{\boldsymbol{X}}(y - x_2, x_2)\, \mathrm{d}x_2. \end{aligned} \tag{F.26}$$

Dies entspricht der charakteristischen Funktion

$$G_Y(k) = G_{X_1,X_2}(k,k). \tag{F.27}$$

Wenn X_1 und X_2 stochastisch unabhängig sind, wird Gl. (F.26) mithilfe von Gl. (F.23) zu

$$p_Y(y) = \int p_{X_1}(x_1)\, p_{X_2}(y - x_1)\, \mathrm{d}x_1,$$

d.h. zur Faltung von p_{X_1} und p_{X_2}. In diesem Fall ist die Varianz von Y

$$\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2. \tag{F.28}$$

[18] Dabei wird angenommen, dass das Produkt definiert werden kann, falls die Funktionen nicht reell- oder komplexwertig sind.

Diese Eigenschaft lässt sich auf alle Kumulanten von Y verallgemeinern (aber nicht auf die zentralen Momenten!), was aus der Gleichheit

$$G_Y(k) = G_{X_1}(k)\, G_{X_2}(k)$$

folgt.

F.4.4 Mehrdimensionale Gauß-Verteilung

Seien Σ eine positiv definite symmetrische $D \times D$ Matrix und $\boldsymbol{\mu}$ ein D-dimensionaler Vektor. Die multivariate Gauß-Verteilung für Zufallsvariablen $\boldsymbol{X} = (X_1, \ldots, X_D)$ ist

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^D \det \Sigma}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right], \tag{F.29}$$

wobei der Index $^{\mathsf{T}}$ die Transponierung bezeichnet.

Dann ist $\boldsymbol{\mu}$ der Erwartungsvektor $\langle \boldsymbol{X} \rangle$ und Σ die *Kovarianzmatrix*, deren (i,j)-Element die Kovarianz $\mathrm{Cov}\,(X_i, X_j)$ der Komponenten X_i und X_j ist.

F.5 Zentraler Grenzwertsatz

Sei $(X_1, X_2, \ldots, X_n, \ldots)$ eine Folge von stochastisch unabhängigen eindimensionalen Zufallsvariablen mit derselben Ergebnismenge Ω und derselben Wahrscheinlichkeitsverteilung. Es wird angenommen, dass der Erwartungswert μ und die Varianz σ^2 der Verteilung existieren. Sei

$$Z_N \equiv \frac{1}{N}\sum_{n=1}^{N} X_n \tag{F.30a}$$

die N-te Teilsumme dieser Zufallsvariablen, multipliziert mit einem geeigneten Normierungsfaktor. Dank den Ergebnissen des § F.4.3 ist der Erwartungswert von Z_N genau gleich μ und deren Varianz σ^2/N.

Laut dem *Zentralen Grenzwertsatz* konvergiert die Wahrscheinlichkeitsverteilung von Z_N für $N \to \infty$ gegen die Gauß-Verteilung mit diesen Parametern, d.h. für jedes reelle z

$$p_{Z_N}(z) \underset{N\gg 1}{\sim} \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2/N}\right]. \tag{F.30b}$$

Dieser Satz liegt der wichtigen Rolle der Gauß-Verteilung zugrunde, und wird außer der Mathematik oft als „Gesetz der großen Zahlen“ bezeichnet. Da die Varianz der Verteilung von Z_N mit wachsender N kleiner wird — die relative „Fluktuation“ ist der Ordnung $1\sqrt{N}$ —, weichen die Realisationen z weniger vom Erwartungswert μ ab: die Verteilung nähert sich einer δ-Distribution im Punkt μ an.

Bemerkungen:

∗ Bei der Konvergenz in Gl. (F.30b) handelt es sich um punktweise Konvergenz.

∗ Es existieren weitere ähnliche Sätze — hier wurde nur der klassische Satz von Lindeberg[(ar)]–Lévy[(as)] angegeben — für unabhängige Zufallsvariablen mit unterschiedlichen Verteilungen, für nicht unabhängige Zufallsvariablen...

Elemente eines Beweises:
Die Wahrscheinlichkeitsdichte für Z_N folgt aus den Verallgemeinerungen von Gl. (F.26) und (F.23)

$$p_{Z_N}(z) = \int p_{X_1}(x_1)\cdots p_{X_N}(x_N)\,\delta\left(z - \frac{1}{N}\sum_{n=1}^{N} x_n\right) \mathrm{d}x_1 \cdots \mathrm{d}x_N,$$

(ar) J. W. Lindeberg, 1876–1932 (as) P. Lévy, 1886–1971 wobei $p_{X_1}, \dots, p_{X_N}$ tatsächlich immer die gleiche Dichte p_X ist. Mithilfe der Fourier-Darstellung der δ-Distribution ergibt sich

$$\begin{aligned} p_{Z_N}(z) &= \int p_{X_1}(x_1) \cdots p_{X_N}(x_N) \exp\left[\mathrm{i}k\left(\frac{1}{N}\sum_{n=1}^{N} x_n - z\right)\right] \mathrm{d}x_1 \cdots \mathrm{d}x_N \, \frac{\mathrm{d}k}{2\pi} \\ &= \int \mathrm{e}^{-\mathrm{i}kz} \prod_{n=1}^{N} \left(\int p_X(x_n)\, \mathrm{e}^{\mathrm{i}kx_n/N}\, \mathrm{d}x_n\right) \frac{\mathrm{d}k}{2\pi} = \int \mathrm{e}^{-\mathrm{i}kz} \left[G_X\left(\frac{k}{N}\right)\right]^N \frac{\mathrm{d}k}{2\pi}, \end{aligned}$$

wobei die charakteristische Funktion (F.7a) der Wahrscheinlichkeitsdichte p_X benutzt wurde. Mithilfe der (per Annahme existierende) Taylor-Entwicklung der Letzteren in $k = 0$ erhält man

$$G_X\left(\frac{k}{N}\right) = 1 + \frac{\mathrm{i}k\mu}{N} - \frac{k^2\langle X^2\rangle}{2N^2} + \mathcal{O}\left(\frac{1}{N^3}\right),$$

d.h.

$$N \ln G_X\left(\frac{k}{N}\right) = \mathrm{i}k\mu - \frac{k^2\sigma^2}{2N} + \mathcal{O}\left(\frac{1}{N^2}\right).$$

Daraus folgt schließlich

$$p_{Z_N}(z) \underset{N\gg 1}{\sim} \int \exp\left[-\frac{k^2\sigma^2}{2N} - \mathrm{i}k(z-\mu)\right] \frac{\mathrm{d}k}{2\pi} = \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2/N}\right]. \qquad \square$$

Literatur

- Dieser Kapitel ist stark inspiriert durch van Kampen, *Stochastic Processes in Physics and Chemistry* [22], Kap. I.