

A Grundlagen der Stochastik

Satz A.1 (Axiome der Wahrscheinlichkeit). Folgende Axiome der Wahrscheinlichkeit können definiert werden:

- (1) Die Wahrscheinlichkeit $P(A)$ eines Ergebnisses A bei einem Experiment ist eine eindeutig bestimmte reelle nichtnegative Zahl, die höchstens gleich 1 werden kann, es gilt also

$$0 \leq P(A) \leq 1 . \quad (\text{A.1})$$

- (2) Für ein sicheres Ereignis A eines Experiments gilt

$$P(A) = 1 . \quad (\text{A.2})$$

Für äquivalente Ereignisse B und C bei einem Experiment gilt

$$P(B) = P(C) . \quad (\text{A.3})$$

- (3) Schließen sich zwei Ereignisse B und C bei einem Experiment gegenseitig aus, dann gilt

$$P(B + C) = P(B) + P(C) . \quad (\text{A.4})$$

Definition A.1 (Zufallsvariable, stochastische Variable). Eine Funktion X heißt Zufallsvariable oder stochastische Variable, wenn sie einem Zufallsexperiment zugeordnet ist und folgende Eigenschaften besitzt:

- (1) Die Werte von X sind reelle Zahlen und
- (2) für jede Zahl a und jedes Intervall I auf der Zahlengeraden ist die Wahrscheinlichkeit des Ereignisses “ X hat den Wert a ” oder “ X liegt im Intervall I ” im Einklang mit den Axiomen der Wahrscheinlichkeit.

Unter einem Zufallsexperiment versteht man in diesem Zusammenhang ein Experiment, bei dem das Ergebnis einer einzelnen Ausführung jeweils durch eine einzelne Zahl ausgedrückt werden kann.

Wenn X eine *diskrete Zufallsvariable* ist, dann können allen Werten von X , im Weiteren mit x_1, x_2, \dots bezeichnet, die zugehörigen Wahrscheinlichkeiten $p_1 = P(X = x_1)$, $p_2 = P(X = x_2), \dots$ zugeordnet werden. Die Funktion

$$f(x) = \begin{cases} p_j & \text{für } x = x_j \\ 0 & \text{für alle übrigen } x \end{cases} \quad (\text{A.5})$$

bezeichnet man dann als *Wahrscheinlichkeitsfunktion*. Da die Zufallsvariable X immer einen Wert x_j annimmt, muss auch gelten, dass die Summe aller Wahrscheinlichkeiten

$$\sum_j f(x_j) = 1 \quad (\text{A.6})$$

ist. Die Wahrscheinlichkeit, dass die Zufallsvariable X im Intervall $a < X \leq b$ liegt, errechnet sich dann auf einfache Weise in der Form

$$P(a < X \leq b) = \sum_{a < x_j \leq b} f(x_j) . \quad (\text{A.7})$$

Trägt man die Wahrscheinlichkeit $P(X \leq x)$ als Funktion von x auf, also die Wahrscheinlichkeit eines Experimentes, dessen Ergebnisse $X \leq x$ sind, dann erhält man die *Wahrscheinlichkeitsverteilungsfunktion*

$$F(x) = P(X \leq x) = \sum_{x_j \leq x} f(x_j) . \quad (\text{A.8})$$

Bei einer *stetigen Zufallsvariablen* X lässt sich die Wahrscheinlichkeitsverteilungsfunktion in Integralform

$$F(x) = \int_{-\infty}^x f(v) dv \quad \text{mit} \quad F(\infty) = 1 \quad (\text{A.9})$$

darstellen, wobei der Integrand $f(v)$ eine nichtnegative und bis auf endlich viele Punkte stetige Funktion ist, die man auch als *Wahrscheinlichkeitsdichtefunktion* bezeichnet.

Aufgabe A.1. Zeigen Sie, dass für die stetige Zufallsvariable X die Beziehung

$$P(a < X \leq b) = \int_a^b f(v) dv = F(b) - F(a)$$

gilt.

Abbildung A.1 zeigt den typischen Verlauf der Wahrscheinlichkeitsverteilungsfunktion einer diskreten und einer stetigen Zufallsvariablen.

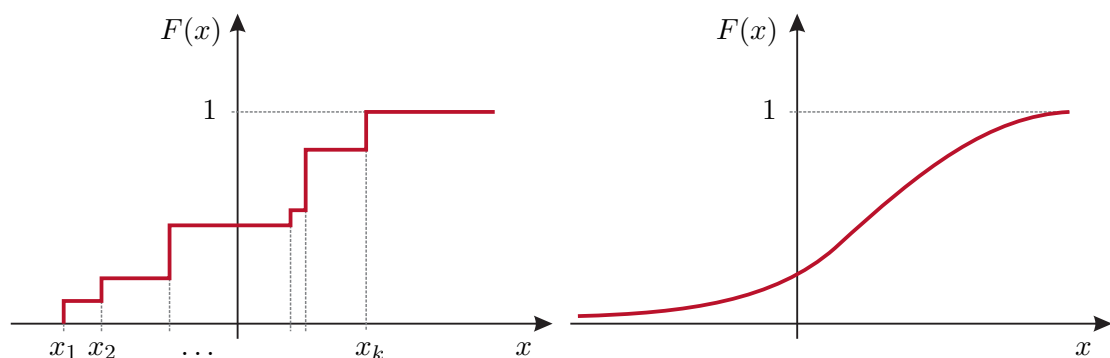


Abbildung A.1: Zur Wahrscheinlichkeitsverteilungsfunktion.

Definition A.2 (Erwartungswert). Unter dem *Erwartungswert* $E(X)$ einer Zufallsvariablen X (auch Mittelwert oder 1. Moment genannt) versteht man im Falle einer diskreten Verteilung

$$E(X) = \sum_j x_j f(x_j) \quad (\text{A.10})$$

und im Falle einer stetigen Verteilung

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx . \quad (\text{A.11})$$

Für eine Funktion $g(X)$ der Zufallsvariablen X gilt allgemein, dass sich der Erwartungswert der Funktion $g(X)$ in der Form

$$E(g(X)) = \sum_j g(x_j) f(x_j) \quad \text{bzw.} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (\text{A.12})$$

errechnet.

Aufgabe A.2. Zeigen Sie die Gültigkeit nachfolgender wichtiger Formel

$$E(\alpha g(X) + \beta h(X)) = \alpha E(g(X)) + \beta E(h(X))$$

mit den Konstanten α und β . Zeigen Sie, dass damit auch

$$E(E(X)) = E(X) \quad \text{und} \quad E(X E(X)) = E(X)^2$$

gilt.

Definition A.3 (Varianz). Die *Varianz* σ_X^2 einer Zufallsvariablen X misst die quadratische Abweichung vom Erwartungswert und ist in der Form

$$\sigma_X^2 = E([X - E(X)]^2) = E(X^2 - 2X E(X) + E(X)^2) = E(X^2) - E(X)^2 \quad (\text{A.13})$$

definiert. Die Varianz wird auch als zweites zentrales Moment und deren positive Quadratwurzel σ_X als Standardabweichung bezeichnet.

Die bisherigen Betrachtungen lassen sich nun einfach auf *mehrere Zufallsvariablen* übertragen. Bei zwei Zufallsvariablen X und Y nimmt die Wahrscheinlichkeitsverteilungsfunktion im diskreten Fall die Form

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_k \leq x} \sum_{y_j \leq y} f(x_k, y_j) \quad (\text{A.14})$$

mit

$$f(x, y) = \begin{cases} p_{kj} & \text{für } x = x_k, y = y_j \\ 0 & \text{für alle übrigen } (x, y) \end{cases} \quad \text{mit} \quad \sum_k \sum_j f(x_k, y_j) = 1 \quad (\text{A.15})$$

an und im stetigen Fall gilt

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(v, w) dv dw \quad \text{mit} \quad F(\infty, \infty) = 1. \quad (\text{A.16})$$

Die so genannten *Randverteilungen* sind durch nachfolgende Ausdrücke

$$F_1(x) = P(X \leq x, Y \text{ beliebig}) = \sum_{x_k \leq x} \underbrace{\sum_j f(x_k, y_j)}_{f_1(x_k)} \quad (\text{A.17})$$

bzw.

$$F_1(x) = P(X \leq x, Y \text{ beliebig}) = \int_{-\infty}^x \underbrace{\int_{-\infty}^{\infty} f(v, w) dv}_{f_1(w)} dw \quad (\text{A.18})$$

und

$$F_2(y) = P(X \text{ beliebig}, Y \leq y) = \sum_{y_j \leq y} \underbrace{\sum_k f(x_k, y_j)}_{f_2(y_j)} \quad (\text{A.19})$$

bzw.

$$F_2(y) = P(X \text{ beliebig}, Y \leq y) = \int_{-\infty}^y \underbrace{\int_{-\infty}^{\infty} f(v, w) dw}_{f_2(v)} dv \quad (\text{A.20})$$

gegeben.

Definition A.4 (Unabhängigkeit von Zufallsvariablen). Zwei Zufallsvariablen X und Y sind genau dann unabhängig, wenn für jedes Paar von Ereignissen der Form

$$a_1 < X \leq b_1 \quad \text{und} \quad a_2 < Y \leq b_2 \quad (\text{A.21})$$

die Beziehung

$$P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = P(a_1 < X \leq b_1) P(a_2 < Y \leq b_2) \quad (\text{A.22})$$

und damit auch

$$f(x, y) = f_1(x)f_2(y) \quad \text{sowie} \quad F(x, y) = F_1(x)F_2(y) \quad (\text{A.23})$$

mit $f_1(x)$, $f_2(y)$, $F_1(x)$ und $F_2(y)$ nach (A.17)–(A.20) gilt.

Satz A.2 (Erwartungswert und Kovarianz zweier Zufallsvariablen). Für den Erwartungswert einer Funktion $g(X, Y)$ in den zwei Zufallsvariablen X und Y gilt

$$E(g(X, Y)) = \sum_k \sum_j g(x_k, y_j) f(x_k, y_j) \quad (\text{A.24})$$

bzw.

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy . \quad (\text{A.25})$$

Der Erwartungswert $E(X + Y)$ einer Summe von Zufallsvariablen X und Y errechnet sich zu

$$E(X + Y) = E(X) + E(Y) . \quad (\text{A.26})$$

Sind zwei Zufallsvariablen X und Y unabhängig, dann gilt die Beziehung (hier für den diskreten Fall mit Hilfe von (A.19) gezeigt)

$$E(XY) = \sum_k \sum_j x_k y_j f(x_k, y_j) = \sum_k \sum_j x_k y_j f_1(x_k) f_2(y_j) = E(X) E(Y) . \quad (\text{A.27})$$

Wenn eine Zufallsvariable Z sich als Summe zweier Zufallsvariablen X und Y ergibt, dann folgt nach (A.13) für die Varianz von Z

$$\begin{aligned} \sigma_Z^2 &= E(Z^2) - E(Z)^2 = E(X^2) + 2E(XY) + E(Y^2) \\ &\quad - E(X)^2 - 2E(X)E(Y) - E(Y)^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} \end{aligned} \quad (\text{A.28})$$

mit der so genannten Kovarianz

$$\sigma_{XY} = E(XY) - E(X)E(Y) . \quad (\text{A.29})$$

Man überzeugt sich leicht, dass (A.29) auch in der Form

$$\sigma_{XY} = E([X - E(X)][Y - E(Y)]) \quad (\text{A.30})$$

angeschrieben werden kann. Sind die Zufallsvariablen X und Y unabhängig, so gilt nach (A.27) $\sigma_{XY} = 0$.

Definition A.5 (Korrelationskoeffizient und Kovarianzmatrix). Der Quotient

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (\text{A.31})$$

wird als *Korrelationskoeffizient* zwischen den Zufallsvariablen X und Y bezeichnet. Ist $r = 0$, dann sind X und Y *unkorreliert*. Man erkennt auch, dass zwei unabhängige Zufallsvariablen X und Y wegen $\sigma_{XY} = 0$ auch unkorreliert sein müssen. Der Korrelationskoeffizient $-1 \leq r \leq 1$ gibt ein Maß für die *lineare Abhängigkeit* von X und Y an.

Für eine vektorwertige Zufallsvariable $\mathbf{X}^T = [X_1 \ X_2 \ \dots \ X_n]$ gilt nun

$$\mathbf{E}(\mathbf{X}^T) = [\mathbf{E}(X_1) \ \mathbf{E}(X_2) \ \dots \ \mathbf{E}(X_n)] \quad (\text{A.32})$$

und unter der *Kovarianzmatrix* der vektorwertigen Zufallsvariablen \mathbf{X} versteht man die Matrix

$$\text{cov}(\mathbf{X}) = \mathbf{E}([\mathbf{X} - \mathbf{E}(\mathbf{X})][\mathbf{X} - \mathbf{E}(\mathbf{X})]^T) = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix}. \quad (\text{A.33})$$

Man erkennt, dass die Kovarianzmatrix symmetrisch ist.

Aufgabe A.3. Zeigen Sie, dass die Beziehung

$$\mathbf{E}(\|\mathbf{X} - \mathbf{E}(\mathbf{X})\|_2^2) = \mathbf{E}([\mathbf{X} - \mathbf{E}(\mathbf{X})]^T [\mathbf{X} - \mathbf{E}(\mathbf{X})]) = \text{spur}(\text{cov}(\mathbf{X}))$$

mit $\text{spur}(\mathbf{S}) = \sum_i s_{ii}$ gilt.

Aus dem bisher Gesagten folgt, dass die Kovarianzmatrix einer vektorwertigen Zufallsvariablen \mathbf{X} , deren Komponenten X_i, X_j für $i \neq j = 1, \dots, n$ unkorreliert sind, eine Diagonalmatrix ist.

Für stochastische Zeitsignale, die aus statistisch identischen Signalquellen stammen, existiert nicht nur eine einzige Realisierung $x_1(t)$ sondern eine ganze Familie (Ensemble) von Zufallszeitfunktionen $\{x_j(t)\}$. Dieses Ensemble von Zufallszeitfunktionen bzw. Zufallsfolgen im zeitdiskreten Fall wird als *zeitkontinuierlicher* bzw. *zeitdiskreter stochastischer Prozess* $\mathbf{x}(t)$ bezeichnet. Eine einzelne Realisierung $x_j(t)$ für festes j nennt man auch *Musterfunktion*. Für jeden festen Zeitpunkt t ist $\mathbf{x}(t)$ eine Zufallsvariable mit einer Wahrscheinlichkeitsverteilungsfunktion (siehe (A.8))

$$F(x, t) = \mathbf{P}(\mathbf{x}(t) \leq x). \quad (\text{A.34})$$

Die Wahrscheinlichkeitsdichtefunktion errechnet sich dann gemäß (A.9) zu

$$f(x, t) = \frac{\partial F(x, t)}{\partial x} . \quad (\text{A.35})$$

Definition A.6 (Mittelwert, Auto- und Kreuzkorrelationsfunktion eines stochastischen Prozesses). Der Mittelwert $\eta_x(t)$ eines stochastischen Prozesses $x(t)$ lautet (siehe auch (A.11))

$$\eta_x(t) = E(x(t)) = \int_{-\infty}^{\infty} x f(x, t) dx . \quad (\text{A.36})$$

Unter der *Autokorrelationsfunktion* $\Phi_{xx}(t_1, t_2)$ eines stochastischen Prozesses $x(t)$ versteht man den Erwartungswert des Produktes $x(t_1)x(t_2)$

$$\Phi_{xx}(t_1, t_2) = E(x(t_1)x(t_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f(x_1, t_1) x_2 f(x_2, t_2) dx_1 dx_2 . \quad (\text{A.37})$$

Die *Kreuzkorrelationsfunktion* $\Phi_{xy}(t_1, t_2)$ zweier stochastischer Prozesse $x(t)$ und $y(t)$ ist durch die Beziehung

$$\Phi_{xy}(t_1, t_2) = E(x(t_1)y(t_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, t_1) y f(y, t_2) dx dy \quad (\text{A.38})$$

gegeben.

Definition A.7 (Auto- und Kreuzkovarianzfunktion). Die *Autokovarianzfunktion* $C_{xx}(t_1, t_2)$ eines stochastischen Prozesses $x(t)$ lautet

$$C_{xx}(t_1, t_2) = E([x(t_1) - \eta_x(t_1)][x(t_2) - \eta_x(t_2)]) = \Phi_{xx}(t_1, t_2) - \eta_x(t_1)\eta_x(t_2) . \quad (\text{A.39})$$

Die *Kreuzkovarianzfunktion* $C_{xy}(t_1, t_2)$ zweier stochastischer Prozesse $x(t)$ und $y(t)$ ergibt sich analog zu

$$C_{xy}(t_1, t_2) = E([x(t_1) - \eta_x(t_1)][y(t_2) - \eta_y(t_2)]) = \Phi_{xy}(t_1, t_2) - \eta_x(t_1)\eta_y(t_2) . \quad (\text{A.40})$$

Aufgabe A.4. Zeigen Sie die Gültigkeit der rechten Identität von (A.39).

Definition A.8 (Stationärer stochastischer Prozess). Man bezeichnet einen stochastischen Prozess $x(t)$ als *stationär im strengen Sinne*, wenn die statistischen Eigenschaften invariant gegenüber Zeitverschiebungen sind, d. h. $x(t)$ und $x(t+c)$ haben für alle c die identischen statistischen Eigenschaften. Der stochastische Prozess $x(t)$ ist *stationär im weiteren Sinne*, wenn der Mittelwert $\eta_x(t) = \eta_x$ konstant ist und die Autokorrelationsfunktion nur von der Zeitdifferenz $\tau = t_2 - t_1$ abhängt, also gilt $\Phi_{xx}(\tau) = E(x(t)x(t+\tau))$.

Die bisher gezeigten Funktionen (A.36)–(A.40) berücksichtigen sämtliche Realisierungen, also das gesamte Ensemble, eines stochastischen Prozesses. Nach der so genannten

Ergoden-Hypothese kann man die Funktionen (A.36)–(A.40) auch mit Hilfe einer einzigen Musterfunktion anschreiben, falls *unendlich lange Zeitabschnitte* betrachtet werden. Ergodische Prozesse sind auch stationär, die Umkehrung gilt im Allgemeinen aber nicht.

Damit ergeben sich die nachfolgende Ausdrücke, jeweils links für den zeitkontinuierlichen und rechts für den zeitdiskreten Fall:

(1) Mittelwert

$$\eta_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt \quad \text{bzw.} \quad \eta_x = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k \quad (\text{A.41})$$

(2) Autokorrelationsfunktion

$$\Phi_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau) dt \quad \text{bzw.} \quad \Phi_{xx}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k x_{k+\tau} \quad (\text{A.42})$$

(3) Kreuzkorrelationsfunktion

$$\Phi_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau) dt \quad \text{bzw.} \quad \Phi_{xy}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k y_{k+\tau} \quad (\text{A.43})$$

Aufgabe A.5. Zeigen Sie, dass $\Phi_{xx}(\tau)$ folgende Eigenschaften erfüllt:

1. $\Phi_{xx}(\tau) = \Phi_{xx}(-\tau)$
2. $\Phi_{xx}(0) = \text{E}(x(t)^2)$
3. $\Phi_{xx}(\infty) = \text{E}(x(t))^2$
4. $\Phi_{xx}(\tau) \leq \Phi_{xx}(0)$

Aufgabe A.6. Zeigen Sie, dass $\Phi_{xy}(\tau)$ folgende Eigenschaften erfüllt:

1. $\Phi_{xy}(\tau) = \Phi_{yx}(-\tau)$
2. $\Phi_{xy}(0) = \text{E}(x(t)y(t))$
3. $\Phi_{xy}(\infty) = \text{E}(x(t))\text{E}(y(t))$
4. $\Phi_{xy}(\tau) \leq \frac{1}{2}[\Phi_{xx}(0) + \Phi_{yy}(0)]$

Aufgabe A.7. Wie lauten im ergodischen Fall die Ausdrücke für die Autovarianz- und Autokovarianzfunktion?

Definition A.9 (Weißes Rauschen). Ein stochastischer Prozess $v(t)$ wird als (*striktes*) *weißes Rauschen* bezeichnet, wenn für alle Zeitpunkte $t_1 \neq t_2$ $v(t_1)$ und $v(t_2)$ statistisch unabhängig sind bzw.

$$C_{vv}(\tau) = C_0 \delta(\tau) \quad \text{mit} \quad \delta(\tau) = \begin{cases} 1 & \text{für } \tau = 0 \\ 0 & \text{sonst} \end{cases} \quad \text{und } C_0 > 0. \quad (\text{A.44})$$

gilt. Es wird weiterhin angenommen, dass der Mittelwert $\eta_v = 0$ ist.

Hinweis: Man kann zeigen, dass im zeitkontinuierlichen Fall die mittlere Leistung des weißen Rauschens unendlich ist und somit dieser Prozess nicht ideal realisiert werden kann. Im Gegensatz dazu bleibt im Zeitdiskreten, wo die Signalfolgenwerte im endlichen Abstand zueinander stehen, die Leistung endlich, womit ideales weißes Rauschen auch realisierbar ist.

A.1 Literatur

- [A.1] R. Isermann, *Identifikation dynamischer Systeme 1 und 2*, 2. Aufl. Berlin, Deutschland: Springer, 1992.
- [A.2] E. Kreyszig, *Statistische Methoden und ihre Anwendungen*, 7. Aufl. Göttingen, Deutschland: Vandenhoeck & Ruprecht, 1998.
- [A.3] D. Luenberger, *Optimization by Vector Space Methods*. New York, USA: John Wiley & Sons, 1969.
- [A.4] A. Papoulis und S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4. Aufl. New York, USA: McGraw-Hill, 2002.