Kapitel 7

Erwartungswert und Integral

Für diskret verteilte Zufallsgrößen haben wir Erwartungswerte in Kapitel vier kennengelernt. Der Begriff des Erwartungswertes war auch Grundlage für die Definition der Varianz einer Zufallsgröße, deren Momente sowie der Kovarianz zweier Zufallsgrößen.
Um sich von der Voraussetzung zu lösen, dass die zugrundeliegenden Zufallsgrößen diskret verteilt sind, erweitern wir den Begriff des Erwartungswertes auf eine möglichst große Klasse von Zufallsgrößen. Das gelingt mit Hilfe der Maß- und Integrationstheorie und soll in diesem Kapitel geschehen. Wir verzichten hier weitgehend auf Beweise, die Darstellung dient nur der Festlegung der Terminologie und der Vorstellung derjenigen Teile der Maß- und Integrationstheorie, die im Rahmen dieser Vorlesung benötigt werden. Für ausführlichere Darstellungen siehe die Vorlesung "Maßtheorie" sowie die Bücher von Bauer (1990), Jacod, Protter (2000) oder Siraev (1988).

7.1 Definitionen

Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum.

Einfache Zufallsgrößen

Definition 7.1 *Eine reellwertige Zufallsgröße X über $(\Omega, \mathfrak{A}, P)$ heißt* einfach *(in der Maßtheorie:* Elementarfunktion*) falls gilt*

$$X(\omega) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega \tag{7.1}$$

für gewisse $n \geq 1, a_i \in R_1, A_i \in \mathfrak{A},\ i = 1, 2, \ldots, n.$

Die Darstellung (7.1) ist nicht eindeutig, da die a_i nicht notwendig verschieden und die A_i nicht notwendig disjunkt sind. Man kann jedoch immer eine Darstellung finden mit $a_i \neq a_j$ und $A_i \cap A_j = \emptyset$ für $i \neq j$.
Jede einfache Zufallsgröße ist diskret verteilt mit der Menge der möglichen Werte $\{a_i, i = 1, \ldots, n\}$ und den Einzelwahrscheinlichkeiten $P(X = a_i)$.

1. Etappe: Erwartungswert einfacher Zufallsgrößen

Definition 7.2 *Es sei X eine einfache Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ der Form (7.1). Als* Erwartungswert *von X oder als* Integral *über X bezüglich P bezeichnet man die Zahl*

$$EX := \sum_{i=1}^{n} a_i P(X = a_i) \tag{7.2}$$

Für EX schreibt man auch $\int\limits_{\Omega} X(\omega) P(d\omega)$ oder kurz $\int\limits_{\Omega} X dP$.

Der Erwartungswert EX hängt nicht von der Darstellung (7.1) ab. Genauer, gelten (7.1) und $X(\omega) = \sum\limits_{j=1}^{m} b_j \mathbb{1}_{B_j}(\omega), \omega \in \Omega$, so haben wir

$$EX = \sum_{i=1}^{n} a_i P(X = a_i) = \sum_{j=1}^{n} b_j P(X = b_j).$$

Offenbar gelten

$$E\mathbb{1}_A = P(A), \quad A \in \mathfrak{A} \quad \text{und} \quad E\mathbb{1} = 1. \tag{7.3}$$

Die hier gegebene Definition stimmt mit der im Abschnitt 4.3. eingeführten Definition des Erwartungswertes diskret verteilter Zufallsgrößen überein. Einfache Zufallsgrößen sind diskret verteilt.

Die Menge aller einfachen Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$ bildet einen linearen Raum, d.h. mit X und Y sind auch alle Linearkombinationen $\alpha X + \beta Y$ $(\alpha, \beta \in R_1)$ einfache Zufallsgrößen.
Die Erwartungswertbildung ist eine *lineare Operation* auf diesem Raum, m.a.W. es gilt

$$E(\alpha X + \beta Y) = \alpha EX + \beta EY. \tag{7.4}$$

Außerdem ist die Erwartungswertbildung eine *monotone Operation*. Sind nämlich X und Y einfache Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$, so gilt

$$X(\omega) \leq Y(\omega),\ \omega \in \Omega \Rightarrow EX \leq EY. \tag{7.5}$$

Zum Beweis von (7.4) und (7.5) wählt man eine Zerlegung $\{C_i, i = 1, \ldots, n\}$ von Ω in Teilmengen C_i aus $\mathfrak{A}$ mit $X = \sum\limits_1^n a_i \mathbb{1}_{C_i}$ und $Y = \sum_1^n b_i \mathbb{1}_{C_i}$.

Nichtnegative Zufallsgrößen

Im nächsten Schritt werden wir den Begriff des Erwartungswertes auf nichtnegative Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$ erweitern. Dazu verwenden wir folgendes Lemma.

Lemma 7.3 *Ist X eine nichtnegative Zufallsgröße über $(\Omega, \mathfrak{A}, P)$, so gibt es eine Folge (X_n) einfacher Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$ mit*

$$\left.\begin{array}{ll} 0 \leq X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega), & \omega \in \Omega, n \geq 1 \\ \lim\limits_{n\to\infty} X_n(\omega) = X(\omega), & \omega \in \Omega. \end{array}\right\} \tag{7.6}$$

Beweis:
Man wähle für jedes $n \geq 1$ und jedes $\omega \in \Omega$

$$X_n(\omega) := \begin{cases} k \cdot 2^{-n}, & \text{falls } X(\omega) \in [k \cdot 2^{-n}, (k+1)2^{-n}) \\ & \text{und } 0 \leq k \leq n2^n - 1 \\ n, & \text{falls } X(\omega) \geq n. \end{cases}$$

Für jedes $n \geq 1$ ist X_n eine Zufallsgröße, also Borel-messbar, da X es ist. Nunmehr ist (7.6) offensichtlich. □

Jede Folge (X_n) mit der Eigenschaft (7.6) nennen wir eine *die nichtnegative Zufallsgröße X approximierende Folge einfacher Zufallsgrößen.*

2. Etappe: Erwartungswert nichtnegativer Zufallsgrößen

Definition 7.4 *Es seien X eine nichtnegative Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ und (X_n) eine X approximierende Folge einfacher Zufallsgrößen. Als Erwartungswert EX von X bezeichnen wir die Zahl*

$$EX := \lim_{n\to\infty} EX_n. \tag{7.7}$$

Der Erwartungswert EX existiert folglich für jede nichtnegative Zufallsgröße X und ist eventuell gleich Unendlich.

Aussage 7.5 *Sind (X_n) und (X_n') zwei die nichtnegative Zufallsgröße X approximierende Folgen, so gilt*

$$\lim_{n\to\infty} EX_n = \lim_{n\to\infty} EX_n' = EX$$

Für den Erwartungswert EX nichtnegativer Zufallsgrößen X gelten die Linearitätseigenschaft (7.4) (zumindest für $\alpha, \beta \geq 0$) und die Monotonieeigenschaft (7.5) sinngemäß.

3. Etappe: Erwartungswert reellwertiger Zufallsgrößen

Im dritten und letzten Schritt erweitern wir den Erwartungswertbegriff auf reellwertige Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$.

Ist X irgend eine solche Zufallsgröße, so zerlegt man sie durch

$$X = X^+ - X^-$$
$$\text{mit } X^+(\omega) := \max(X(\omega), 0) \text{ und } \quad X^-(\omega) := -\min(X(\omega), 0), \omega \in \Omega,$$

in zwei nichtnegative Zufallsgrößen X^+ und X^-. Wir bemerken, dass mit

$|X|(\omega) := |X(\omega)|, \omega \in \Omega$, außerdem die Gleichung $|X| = X^+ + X^-$ gilt.

Definition 7.6 *Man sagt, die Zufallsgröße X über $(\Omega, \mathfrak{A}, P)$ hat einen endlichen Erwartungswert EX, falls $EX^+ < \infty$ und $EX^- < \infty$ gelten. Der Erwartungswert EX wird in diesem Fall definiert als $EX := EX^+ - EX^-$.*

Gilt $EX^+ < \infty$ oder $EX^- < \infty$ so sagt man, X besitze einen Erwartungswert und setzt ebenfalls $EX = EX^+ - EX^-$. In diesem Fall kann $EX = \infty$ bzw. $= -\infty$ gelten. Ist $EX^+ = EX^- = \infty$, so heißt es, X habe keinen Erwarungswert.

In anderer Sprechweise sagt man, falls X einen endlichen Erwartungswert hat, X sei bezüglich P integrierbar und schreibt für EX auch

$$\int_\Omega X(\omega)P(d\omega) \text{ oder kurz } \int_\Omega X dP.$$

Der Erwartungswert EX von X wird in diesem Zusammenhang auch als *Integral über X* bez. P, kurz P-Integral über X, bezeichnet.

Existiert EX, so existiert für jedes $A \in \mathfrak{A}$ auch $E(X\mathbb{1}_A)$, wir schreiben dafür auch $\int_A X dP$.

Gilt für eine nichtnegative Zufallsgröße X die Gleichung $EX = 0$, so folgt $P(X = 0) = 1$, die Zufallsgröße X hat also eine "entartete" Verteilung, sie nimmt mit Wahrscheinlichkeit Eins den Wert Null an. Ein Beispiel dafür haben wir bei der Einführung der gleichmäßigen Verteilung auf $[0, 1)$ gesehen. Für

$X = \mathbb{1}_Q$ mit Q = Menge der rationalen Zahlen aus $[0,1)$ gilt $EX = \lambda_{[0,1)}(Q) = 0$. Die Abbildung X ist deswegen aber nicht identisch Null, sondern nur P-fast sicher gleich Null.

Insbesondere folgt für jede Zufallsgröße X mit $E|X| = 0$ die Eigenschaft $P(X = 0) = 1$.

P-Äquivalenzklassen von Zufallsgrößen

Definition 7.7 *Zwei Zufallsgrößen X und Y über $(\Omega, \mathfrak{A}, P)$ heißen* P-äquivalent *oder einfach* äquivalent *, falls gilt*

$P(\{\omega \in \Omega | X(\omega) \neq Y(\omega)\}) = 0.$
Alle zueinander P-äquivalenten Zufallsgrößen fasst man zu einer Äquivalenzklasse *zusammen.*

Sind zwei Zufallsgrößen X und Y P-äquivalent und existiert der Erwartungswert EX, so existiert auch EY und beide sind einander gleich. Der Erwartungswert ist also ein Funktional auf der Menge aller Äquivalenzklassen.

7.2 Einige Eigenschaften des Erwartungswertes

Es seien $(\Omega, \mathfrak{A}, P)$ irgendein Wahrscheinlichkeitsraum und $X, Y, \cdots$ reellwertige Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$.

Mit $\mathfrak{L}_1(\Omega, \mathfrak{A}, P)$ bezeichnen wir die Menge aller reellwertigen Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$ mit endlichem Erwartungswert. Wir fassen einige Eigenschaften des Erwartungswertes in folgender Aussage zusammen.

Aussage 7.8

a) $X, Y \in \mathfrak{L}_1 \Longrightarrow \alpha X + \beta Y \in \mathfrak{L}_1$ *und* $E(\alpha X + \beta Y) = \alpha EX + \beta EY, (\alpha, \beta \in R_1)$
($\mathfrak{L}_1$ ist ein linearer Raum und $X \to EX$ ein lineares Funktional auf $\mathfrak{L}_1$)

b) $X \in \mathfrak{L}_1, X \leq Y \;\; \mathrm{P-f.s.} \Longrightarrow \mathrm{EX} \leq \mathrm{EY} \leq \infty$
(insbesondere folgt aus $Y \geq 0 \; \mathrm{P-f.s.}$ *die Ungleichung* $EY \geq 0$*)*

c) $X \in \mathfrak{L}_1 \Longleftrightarrow |X| \in \mathfrak{L}_1$*, in diesem Fall gilt* $|EX| \leq E|X|$

d) Ist X $\mathrm{P-f.s.}$ *beschränkt* ($|X| \leq C$ $\mathrm{P-f.s.}$ *für ein* $C > 0$)*, so besitzt* X *einen endlichen Erwartungswert* EX*.*

Ungleichungen

Im Folgenden stellen wir einige Ungleichungen den Erwartungswert von Zufallsgrößen betreffend zusammen, die in der Wahrscheinlichkeitstheorie relevant sind.

a) *Ungleichung von Tschebychev:*
Ist X eine nichtnegative Zufallsgröße, so gilt für jeden $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon} \tag{7.8}$$

Beweis:

$$EX \geq E(\mathbb{1}_{\{X \geq \varepsilon\}} \cdot X) \geq E\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}} = \varepsilon P(X \geq \varepsilon)$$

□

b) *Ungleichung von Cauchy-Schwarz:*

Ist $E(X^2) < \infty$ und $E(Y^2) < \infty$, dann gilt $E|XY| < \infty$
und

$$(E(XY))^2 \leq E(X^2) \cdot E(Y^2) \tag{7.9}$$

Das Gleichheitszeichen gilt genau dann, wenn $aX + bY = 0$ für gewisse $a, b \in R_1$ $\mathrm{P-f.s.}$ gilt.

Beweis: O.B.d.A. sei $EX^2 > 0, EY^2 > 0$, wir setzen

$$\tilde{X} := \frac{X}{\sqrt{EX^2}}, \tilde{Y} := \frac{Y}{\sqrt{EY^2}}.$$

Wegen $E(\tilde{X} - \tilde{Y})^2 \geq 0, \quad E(\tilde{X} + \tilde{Y})^2 \geq 0$, und $E\tilde{X}^2 = E\tilde{Y}^2 = 1$

gilt

$$-1 \leq E\tilde{X}\tilde{Y} \leq 1$$

mit $|E\tilde{X}\tilde{Y}| = 1$ genau dann, wenn $\tilde{X} = \tilde{Y}$ oder $\tilde{X} = -\tilde{Y}$ $\mathrm{P}-$f.s.

Daraus folgt die Behauptung. □

Bemerkung: Die Ungleichung (7.9) bleibt erhalten, wenn man auf der linken Seite $E|XY|$ an Stelle $E(XY)$ setzt. Der Beweis verläuft analog.

c) *Ungleichung von Jensen:*
Es seien g eine von unten konvexe und Borel-messbare Funktion auf R_1 und X eine reellwertige Zufallsgröße mit $E|X| < \infty$. Dann gilt

$$g(EX) \leq Eg(X) \leq \infty. \tag{7.10}$$

Beweis: Da g von unten konvex ist, gibt es zu jedem $x_0 \in R_1$ eine Zahl $\lambda(x_0)$ mit

$$g(x_0) + (x - x_0)\lambda(x_0) \leq g(x).$$

Wir setzen $x = X, x_0 = EX$ und erhalten damit

$$g(EX) + (X - EX)\lambda(EX) \leq g(X),$$

daraus folgt $g(EX) \leq Eg(X)$. □

Die Jensen'sche Ungleichung impliziert zwei weitere Ungleichungen, die wir hier nur angeben, für einen Beweis siehe z. B. Siraev (1988), Kap. II, § 6.

d) *Hölder-Ungleichung:* Es sei $1 < p < \infty, 1 < q < \infty, \frac{1}{p} + \frac{1}{q} = 1$. Wenn $E|X|^p < \infty, E|Y|^q < \infty$, so ist $E|XY| < \infty$, und es gilt

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}$$

($p = q = 2$: Cauchy-Schwarz-Ungleichung)

e) *Minkovski-Ungleichung:* Wenn $E|X|^p < \infty, E|Y|^p < \infty$, für ein p mit $1 \leq p < \infty$, dann gilt $E|X+Y|^p < \infty$ und

$$(E|X+Y|^p)^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}.$$

Die Räume L^p

Es sei $p \in [1, \infty)$ und $\mathfrak{L}^p(\Omega, \mathfrak{A}, P)$ die Menge aller reellwertigen Zufallsgrößen X über $(\Omega, \mathfrak{A}, P)$ mit $E(|X|^p) < \infty$. Die Menge aller Äquivalenzklassen von Zufallsgrößen X aus $\mathfrak{L}^p$ werde mit $L^p(\Omega, \mathfrak{A}, P)$ bezeichnet (siehe Definition 7.6).

Aussage 7.9 *Es sei* $p \in [1, \infty)$

a) Die Menge $\mathfrak{L}^p = \mathfrak{L}^p(\Omega, \mathfrak{A}, P)$ *ist ein linearer Raum.*

b) $L^p(\Omega, \mathfrak{A}, P)$ *ist mit der Norm*

$$\| X \|_p := (E|X|^p)^{\frac{1}{p}} \quad , \quad X \in L^p$$

ein normierter Raum, sogar ein Banachraum.

c) Es gilt für alle p, p' *mit* $1 \leq p < p' < \infty$

$$L^{p'}(\Omega, \mathfrak{A}, P) \subseteq L^p(\Omega, \mathfrak{A}, P) \text{ und}$$

$$\| X \|_p \leq \| X \|_{p'}, X \in L^{p'}, \text{ (Ungleichung von Ljapunov)} \tag{7.11}$$

Insbesondere gilt

$$E|X| \leq (EX^2)^{\frac{1}{2}}. \tag{7.12}$$

Vertauschung von Grenzwert und Erwartungswert

Es sei $(X_n, n \geq 1)$ eine Folge reellwertiger Zufallsgrößen über $(\Omega, \mathfrak{A}, P)$.

Definition 7.10 *Man sagt, die Folge $(X_n, n \geq 1)$ konvergiert P-fast sicher gegen eine Zufallsgröße X über $(\Omega, \mathfrak{A}, P)$, falls*

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

Die folgenden drei Aussagen betreffen das Verhältnis zwischen Grenzwerten und Erwartungswerten.

Aussage 7.11 *(Satz von der majorisierten Konvergenz) Konvergiert $(X_n, n \geq 1)$ P-fast sicher gegen X und gibt es eine P-integrierbare Zufallsgröße Z mit $|X_n| \leq Z \; P - f.s.$ für alle $n \geq 1$, so ist auch X bezüglich P integrierbar, und es gilt*

$$\lim_{n\to\infty} EX_n = E \lim_{n\to\infty} X_n = EX.$$

Aussage 7.12 *(Satz von der monotonen Konvergenz) Ist $(X_n, n \geq 1)$ eine monoton wachsende Folge P-integrierbarer Zufallsgrößen, so gilt für $X := \lim\limits_{n\to\infty} X_n$ die Beziehung*

$$EX = \lim_{n\to\infty} EX_n \leq \infty.$$

Aussage 7.13 (Lemma von Fatou): *Sind Y und Z zwei P-integrierbare Zufallsgrößen, so gilt*

$$X_n \leq Y \mathrm{P-f.s.} \textit{ für alle } n \geq 1 \Longrightarrow E(\lim_{n\to\infty} \sup X_n) \geq \lim_{n\to\infty} \sup EX_n$$

$$X_n \geq Z \mathrm{P-f.s.} \textit{ für alle } n \geq 1 \Longrightarrow E(\lim_{n\to\infty} \inf X_n) \leq \lim_{n\to\infty} \inf EX_n$$

Die folgende Aussage gestattet es, die Berechnung des Erwartungswertes einer Zufallsgröße auf ein Integral bezüglich ihrer Wahrscheinlichkeitsverteilung zurückzuführen.

Aussage 7.14 (Substitutionsformel): *Es sei X eine Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit Werten in $(E, \mathfrak{E})$ und der Wahrscheinlichkeitsverteilung P^X (siehe ...). Weiterhin sei h eine $\mathfrak{E} - \mathfrak{B}_1-$messbare Abbildung von E in R_1. Dann gilt:*

a) $h(X)$ ist P-integrierbar genau dann, wenn $h(.)$ bezüglich P^X integrierbar ist.

b) Im Falle von a) gilt

$$Eh(X) = \int_\Omega h(X(\omega))P(d\omega) = \int_E h(x)P^X(dx). \tag{7.13}$$

Beweis: Wir gehen zurück auf die Definition von P^X. Es gilt

$$P^X(B) = P(X^{-1}(B)) \ , \quad B \in \mathfrak{E}.$$

Daraus folgt

$$E(\mathbb{1}_B(X)) = P(X^{-1}(B)) = P^X(B) = \int_E \mathbb{1}_B(x)P^X(dx) \tag{7.14}$$

Ist h eine einfache Funktion (Elementarfunktion, endliche Linearkombination aus messbaren Indikatorfunktionen) so folgt aus (7.14) die Eigenschaft (7.13) auf Grund der Linearität der Erwartungswertoperation.
Wenn h nichtnegativ ist, so wählen wir eine h approximierende Folge h_n einfacher Funktionen:

$$0 \leq h_n \leq h_{n+1} \leq h$$
$$\lim_{n\to\infty} h_n(x) = h(x), \quad x \in E.$$

Dann gilt $h_n(X) \uparrow h(X)$ und wegen des Satzes (7.11) von der monotonen Konvergenz (zweimal angewandt)

$$Eh(X) = E \lim h_n(X) = \lim Eh_n(X) = \lim \int_E h_n(x)P^X(dx) = \int_E h(x)P^X(dx)$$

Das beweist a) und b) für nichtnegative h. Für beliebiges h benutzen wir wieder die Zerlegung $h = h^+ - h^-$. □

Varianz, Kovarianz und Korrelation

Wir haben bei diskret verteilten Zufallsgrößen gesehen, dass zur Beurteilung einer Wahrscheinlichkeitsverteilung neben dem Erwartungswert, der den "Schwerpunkt" der Verteilung beschreibt, auch die Varianz oder Streuung von Bedeutung ist. Sie ist eine Maßzahl, wie breit die möglichen Werte der Zufallsgröße um den Erwartungswert (mit ihren Wahrscheinlichkeiten gewichtet) gelagert sind bzw. wie stark Realisierungen einer zugrundeliegenden Zufallsgröße um ihren Mittelwert "streuen".

Der Begriff der Varianz oder der Streuung überträgt sich mit dem nunmehr bereit stehenden Begriff des Erwartungswertes beliebiger Funktionen von Zufallsgrößen problemlos auf unseren allgemeinen Fall.

Definition 7.15 *Für jedes $X \in L_2$ wird durch*

$$D^2(X) = Var(X) := E((X - EX)^2)$$

die Varianz *(oder* die Streuung*) von X definiert. Sie wird häufig auch mit σ_X^2 bzw. einfach mit σ^2 bezeichnet. Die Zahl $\sigma_X = (\sigma_X^2)^{\frac{1}{2}}$ heißt* Standardabweichung *der Zufallsgröße X.*

Es gilt

$$D^2(X) = E((X - EX)^2) = EX^2 - 2EXEX + (EX)^2 = EX^2 - (EX)^2 \tag{7.15}$$

Die Wirkung linearer Transformationen

Hat die Zufallsgröße X einen endlichen Erwartungswert, so gilt für alle $a, b \in R_1$ die Gleichung $E(aX + b) = aEX + b$.

Ist $D^2X < \infty$, so besitzt für jede reelle Zahl a die Zufallsgröße aX die Varianz

$$D^2(aX) = a^2 D^2 X,$$

und für jedes $b \in R_1$ gilt

$$D^2(aX + b) = a^2 D^2 X.$$

Ist $D^2 X > 0$, so bildet

$$X^* := \frac{X - EX}{\sqrt{D^2 X}} \tag{7.16}$$

eine *standardisierte* Zufallsgröße, d. h., es gilt

$$EX^* = 0 \text{ und } D^2 X^* = 1.$$

Bemerkung 7.16 Hat eine Zufallsgröße X eine positive Streuung $D^2 X$ (oder ist diese gleich Unendlich), so handelt es sich um eine echte Zufallsgröße in dem Sinne, dass ihr Wert vor Ausführung des zugrunde liegenden Experimentes unbestimmt ist. Ihre möglichen Werte besitzen eine "echte" Wahrscheinlichkeitsverteilung, die Gesamtwahrscheinlichkeit Eins verteilt sich auf mehrere verschiedene mögliche Werte.

Dagegen gilt $D^2 X = 0$ genau dann, wenn $P(X = EX) = 1$ erfüllt ist, wenn also X mit Wahrscheinlichkeit Eins nur einen einzigen Wert annehmen kann, der dann natürlich der Erwartungswert von X ist.

Aus Formel (7.8) folgt die

Aussage 7.17 (Tschebyschev'sche Ungleichung) *Ist $D^2 X < \infty$, so gilt für jedes $\varepsilon > 0$*

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

Ist die Streuung $D^2 X$ positiv aber klein, so besagt die Tschebyschev'sche Ungleichung, dass die möglichen Werte von X, die weit von EX entfernt liegen, bei einer Realisierung der Zufallsgrß̈e X nur mit sehr kleiner Wahrscheinlichkeit (die aber durchaus positiv ist) auftreten werden.

7.3 Dichten eindimensionaler Verteilungen

In diesem und im folgenden Abschnitt erweitern wir den in Abschnitt 3.5 eingeführten Begriff der Dichte einer Wahrscheinlichkeitsverteilung. Wir stützen uns dabei auf Vorkenntnisse über das Lebesguemaß λ auf $(R_1, \mathscr{B}_1)$ aus der Maßtheorie-Vorlesung.
Die Integration über reellwertige Borel-messbare Funktionen f auf R_1 bez. des Lebesguemaßes definiert man völlig analog zur Definition des Erwartungswertes, d. h. des Integrales bezüglich des Wahrscheinlichkeitsmaßes P in Abschnitt 7.1.

Statt $\int\limits_{R_1} f(x)\lambda(dx)$ schreiben wir $\int\limits_{R_1} f(x)dx$.

Definition 7.18 *Ist Q ein Wahrscheinlichkeitsmaß auf $(R_1, \mathscr{B}_1)$, und existiert eine nichtnegative Borelfunktion f auf R_1, so dass*

$$F_Q(x) \coloneqq Q((-\infty, x]) = \int\limits_{(-\infty,x]} f(y)dy = \int\limits_{R_1} f(y)\mathbb{1}_{(-\infty,x]}(y)dy, x \in R_1 \tag{7.17}$$

gilt, so heißt f die Dichtes des Maßes Q. *Ist $Q = P^X$ für eine Zufallsgröße X, so nennt man f auch die* Dichte der Zufallsgröße X.

Aus (7.17) folgt wie üblich mit Hilfe des Erweiterungssatzes für σ-additive Mengenfunktionen

$$Q(B) = \int\limits_{R_1} f(y)\mathbb{1}_B(y)dy =: \int\limits_B f(y)dy \tag{7.18}$$

für jedes $B \in \mathscr{B}_1$.

Aussage 7.19 *Genau dann besitzt eine Wahrscheinlichkeitsverteilung Q auf $(R_1, \mathfrak{L}_1)$ eine Dichte f wenn ihre Verteilungsfunktion F_Q Lebesgue-fast überall differenzierbar ist. In diesem Fall gilt $\frac{dF_Q}{dx} = f(x)$ Lebesgue-fast überall.*

Der Beweis dieser Aussage ist Gegenstand der Analysis monotoner Funktionen auf R_1, siehe z. B. I.P.Natanson, Theorie der Funktionen einer reellen Veränderlichen, Akademie Verlag, 1961.

Aussage 7.20 *Eine nichtnegative Borelmeßbare Funktion f auf R_1 ist die Dichte einer Wahrscheinlichkeitsverteilung Q auf $\mathscr{B}_1$ genau dann, wenn gilt*

$$\int\limits_{R_1} f(x)dx = 1.$$

Die Verteilung Q ist in diesem Fall durch die Formel in (7.18) gegeben.
Besitzt eine Wahrscheinlichkeitsverteilung Q auf $(R_1, \mathscr{B}_1)$ eine Dichte f, so bestimmt f das Maß Q eindeutig. Andererseits ist für je zwei Dichten f_1 und f_2 von Q

$$\lambda(\{x \in R_1 | f_1(x) \neq f_2(x)\}) = 0,$$

d. h., f_1 und f_2 sind Lebesgue - fast überall gleich.

Beweisskizze:

$$\int\limits_{R_1} f(x)dx = \lim_{y\to\infty} Q((-\infty, y]) = 1.$$

Wenn $f \geq 0$ gegeben ist, so setzt man

$$Q(B) := \int\limits_{R_1} f(x)\mathbb{1}_B(x)dx = \int\limits_{B} f(x)dx.$$

Sind f_1 und f_2 Dichten von Q, so gilt

$$\int\limits_{R_1} \mathbb{1}_{\{f_1<f_2\}}(x)(f_2(x) - f_1(x))dx = 0,$$

folglich ist $\lambda(\{f_1 < f_2\}) = 0$, und somit auch $\lambda(\{f_1 \neq f_2\}) = 0$.

Mit Hilfe der folgenden Aussage gelingt es, Erwartungswerte der Form $Eg(X)$ auf Integrale bezüglich des Lebesguemaßes zurückzuführen.

Aussage 7.21 *Es sei X eine reellwertige Zufallsgröße mit der Dichte f. Ist g eine Borel-messbare Funktion auf R_1, so gilt*

a) $g(.)$ ist bezüglich P^X integrierbar genau dann, wenn $g(.)f(.)$ bezüglich des Lebesguemaßes integrierbar ist,

b) im Fall a) gilt

$$Eg(X) = \int_{R_1} g(x)P^X(dx) = \int_{R_1} g(x)f(x)dx. \tag{7.19}$$

Beweis: Für $g = \mathbb{1}_B$ mit $B \in \mathscr{B}_1$ hat (7.19) die Form $E\mathbb{1}_B(X) = P^X(B) = \int_{R_1} \mathbb{1}_B(x)f(x)dx$.

Diese Gleichung ist aber auf Grund von (7.18) und $E\mathbb{1}_B(X) = P(X \in B)$ richtig, man setze $Q = P^X$.

Wegen der Linearität der Erwartungswertbildung folgt damit (7.19) für alle einfachen Funktionen $g(\cdot)$ (Elementarfunktionen). Für allgemeines nichtnegatives g ergibt sich (7.19) und auch a) aus dem Satz über die monotone Konvergenz.
Der Fall beliebiger Funktionen g folgt wie üblich mittels $g = g^+ - g^-$. $\square$

Lebesgue- und Riemannintegrale

Der Einfachheit und Allgemeinheit der Definition von P-Integralen steht die Kompliziertheit ihrer konkreten Ausrechnung auf der Grundlage ihrer Definition gegenüber. Andererseits verfügt man mit der Theorie des Riemannintegrals und seiner zahlreichen Berechnungsmethoden über ein sehr leistungsfähiges Werkzeug zur Berechnung von Integralen. Wir geben im Folgenden die Beziehungen zwischen beiden Integralarten an und gewinnen damit die Möglichkeit, in vielen Fällen Erwartungswerte, Streuungen und andere Kenngrößen von Verteilungen konkret ausrechnen zu können. Die Beweise findet man in der Literatur zur Maß- und Integrationstheorie, siehe z. B. Elstrodt (1996), Bauer (1992) oder die Vorlesung Maßtheorie.

Aussage 7.22 *Es sei f eine beschränkte Borel-messbare Funktion auf dem endlichen Intervall $[a,b]$. Dann gilt:*

a) f ist L-integrierbar,

b) f ist R-integrierbar genau dann, wenn $\{x \in [a,b] : f \text{ ist unstetig bei } x\}$ das Lebesguemaß Null hat.

Im Fall b) gilt

$$(R) - \int\limits_{[a,b]} f(x)dx = (L) - \int\limits_{[a,b]} f(x)dx.$$

Im Fall eines unendlichen Integrationsbereiches, z. B. $I = (\infty, a], = (a, \infty)$ oder $= R_1$, hat man für Funktionen f, die auf jedem kompakten Intervall $[a,b] \subseteq I$ Riemannintegrierbar sind, den Begriff des *uneigentlichen Riemannintegrals.* Man sagt (hier für $I = R_1$ aufgeschrieben), das uneigentliche Riemannintegral über f existiert, falls der Grenzwert

$$(R) - \int\limits_{-\infty}^{\infty} f(x)dx := \lim_{\substack{a \to -\infty \\ b \to +\infty}} (R) - \int\limits_{[a,b]} f(x)dx,$$

existiert und endlich ist.

Wir vergleichen uneigentliche Riemannintegrale mit Lebesgueintegralen und bemerken als Erstes die folgende

Aussage 7.23 *Ist f eine nichtnegative Funktion auf $I_{[a,\infty)}$, und ist f auf jedem Intervall $[a,b]$ für $b > a$ R-integrierbar, so gilt*

$$\lim_{b \to \infty} (R) - \int\limits_{[a,b]} f(x)dx = (L) - \int\limits_{R_1} f(x)dx.$$

Anders ausgedrückt, das uneigentliche R-Integral über eine nichtnegative Funktion f existiert genau dann, wenn das L-Integral existiert und endlich ist. In diesem Fall sind beide gleich.

Beweis: Die Folge (f_n), definiert durch $f_n := f \cdot \mathbb{1}_{[a,n]}$ konvergiert monoton gegen f. Die Aussagen 7.12 und 7.21 implizierten

$$(L) - \int\limits_{[a,\infty)} f(x)dx = \lim_{n\to\infty} (L) - \int\limits_{[a,\infty)} f_n(x)dx =$$

$$\lim_{n\to\infty} (R) - \int\limits_{[a,n]} f(x)dx = (R) - \int\limits_{[a,\infty)} f(x)dx.$$

□

Wir setzen den Vergleich beider Integralarten fort mit der folgenden Bemerkung:

Das uneigentliche Riemannintegral kann existieren und endlich sein, obwohl f nicht Lebesgueintegrierbar ist.

Beispiel 7.24 Für f, definiert durch $f(x) = \frac{\sin x}{x}$, $x > 0$, gilt

$$(R) - \int\limits_0^\infty f(x)dx = \sum_{k=0}^\infty \int\limits_{[\pi k, \pi(k+1)]} \frac{\sin x}{x}\, dx.$$

Die Reihe konvergiert, da sie alternierend ist und die Reihenglieder gegen Null konvergieren. Das Lebesgueintegral $(L) - \int\limits_{[0,\infty)} f(x)dx$ existiert nicht, da f^+ und f^- kein endliches Lebesgueintegral besitzen.

Die folgende Aussage gibt eine Bedingung an, unter der Erwartungswerte der Form $Eh(X)$ mit Hilfe von Riemannintegralen berechnet werden können.

Aussage 7.25 *Es sei X eine reellwertige Zufallsgröße mit der Dichte f und h eine Funktion von R_1 in sich.*
Sind h und f Lebesgue-fast-überall stetig und ist h nichtnegativ, so gilt

$$Eh(X) = (R) - \int\limits_{R_1} h(x)f(x)dx \tag{7.20}$$

Die Gleichung (7.20) gilt auch für h *mit* $E|h(X)| < \infty$ *oder, äquivalent,*

$$(R) - \int_{R_1} |h(x)| f(x) dx < \infty.$$

Der Beweis ergibt sich unmittelbar aus Aussage 7.22.

Es sei X eine reellwertige Zufallsgröße mit der Dichte f.

Folgerung 7.26 Das n-te Moment $\mu_n := E(X^n)$ der Zufallsgröße X existiert und ist endlich genau dann, wenn $\int_{R_1} x^n f(x) dx$ existiert und endlich ist. In diesem Fall gilt

$$E(X^n) = \int_{R_1} x^n f(x) dx. \tag{7.21}$$

Insbesondere ergibt sich

$$EX = \int_{R_1} x f(x) dx \text{ und } D^2 X = \int_{R_1} (x - EX)^2 f(x) dx. \tag{7.22}$$

Dabei sind die Integrale als Lebesgueintegrale zu verstehen, die unter geeigneten Voraussetzungen (s. oben) auch zu Riemannintegralen werden.

Beispiele für Dichten auf R_1

Obwohl jede nichtnegative Borel-messbare Funktion f mit $\int_{R_1} f(x)dx = 1$ Dichte einer Wahrscheinlichkeitsverteilung auf $(R_1, \mathscr{B}_1)$ ist, sind viele theoretisch und praktisch wichtige Dichten stetig oder stückweise stetig. Wir geben einige davon an. Ihre Bedeutung wird im weiteren Verlauf der Vorlesung noch diskutiert.

Es sei X eine reellwertige Zufallsgröße mit der Dichte f.
Man sagt X besitze eine

a) *gleichmäßige Verteilung auf* $[a,b]$, falls

$$f(x) = \mathbb{1}_{[a,b]}(x)\frac{1}{b-a}\ ,\ x \in R_1.$$

Bezeichnung: $X \sim U([a,b])$

$$EX = \frac{a+b}{2}\ ,\ D^2X = \frac{(b-a)^2}{12}$$

b) *Exponentialverteilung mit dem Parameter* $\lambda(\lambda > 0)$, falls

$$f(x) = \lambda \cdot \mathbb{1}_{[0,\infty)}(x)\exp(-\lambda x),\ x \in R_1.$$

Bezeichnung: $X \sim Exp(\lambda)$

$$EX = \frac{1}{\lambda},\ D^2X = \frac{1}{\lambda^2}$$

c) *Gammaverteilung mit den Parametern* $\alpha, \lambda(\alpha > 0, \lambda > 0)$, falls

$$f(x) = \mathbb{1}_{(0,\infty)}(x)\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda \mathrm{x}}\ ,\ \mathrm{x} \in \mathrm{R}.$$

Bezeichnung: $X \sim \Gamma(\alpha, \lambda)$

$$EX = \frac{\alpha}{\lambda},\ D^2X = \frac{\alpha}{\lambda^2}$$

Für $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ ist diese Verteilung auch als *"χ^2-Verteilung mit n Freiheitsgraden"* ($n \geq 1$) bekannt.

d) *Normal-* oder *Gaußsche Verteilung mit den Parametern μ und σ^2* ($\mu \in R_1, \sigma^2 > 0$),
falls $f(x) = (2\pi\sigma^2)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), x \in R_1$

Bezeichnung: $X \sim N(\mu, \sigma^2)$

$$EX = \mu,\ D^2X = \sigma^2$$

Im Fall $\mu = 0, \sigma^2 = 1$ spricht man von einer *"Standardnormalverteilung"*. Ihre Verteilungsfunktion wird mit Φ bezeichnet:

$$\Phi(x) = P(X \leq x) = \int\limits_{(-\infty,x]} (2\pi)^{\frac{1}{2}} \exp\left(-\frac{y^2}{2}\right) dy\ ,\ x \in R_1.$$

Sie ist nicht explizit berechenbar und ist deshalb vertafelt.

e) *Cauchyverteilung, mit dem Parameter $a \in R_1$* falls

$f(x) = \frac{1}{\pi}\frac{1}{1+(x-a)^2} x \in R_1.$

Erwartungswert und Streuung der Cauchyverteilung existieren nicht.

Transformationssatz für Dichten

Es sei X eine reellwertige Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit der Dichte f. Häufig hat man die Verteilung einer Zufallsgröße Y zu berechnen, die eine Funktion von X ist. Dazu nehmen wir an, h sei eine Borel-messbare Funktion von R_1 in sich, und es gelte

$$Y(\omega) := h(X(\omega)), \quad \omega \in \Omega.$$

Offenbar gilt für die Verteilungsfunktion F_Y von Y

$$F_Y(y) = P(Y \le y) = P(\{\omega \in \Omega | h(X(\omega)) \le y\}) =$$

$$P(\{\omega \in \Omega | X(\omega) \in h^{-1}((-\infty, y])\}) = P(\{\omega \in \Omega | X(\omega) \in \{x : h(x) \le y\}\}) =$$

$$\int\limits_{\{x:h(x)\le y\}} f(s)ds. \tag{7.23}$$

Aus dieser Gleichung gewinnen wir folgende

Aussage 7.27 *Ist f_X eine stetige Dichte von X, $\{x \in R_1, |f(x) > 0\}$ ein Intervall I, und ist h eine stetig differenzierbare, streng monotone Funktion von I in R_1 mit $h'(x) \neq 0$ für alle $x \in I$, gilt $Y = h(X)$ und setzt man $g(y) = h^{-1}(y)$, so besitzt Y ebenfalls eine Dichte f_Y, und es gilt*

$$f_Y(y) = f_X(g(y))|g'(y)|, \quad y \in R_1. \tag{7.24}$$

Beweis: Es sei h monoton wachsend. Dann gilt

$$F_Y(y) = P(Y \le y) = P(h(X) \le y) =$$

$$P(X \le g(y)) = \int\limits_{(-\infty, g(y)]} f_X(x)dx.$$

Darauf folgt, dass F_Y differenzierbar ist, und dass gilt

$$f_Y(y) := F_y'(y) = f_X(g(y)) \cdot g'(y).$$

Ist h monoton fallend, so haben wir $f_Y(y) = f_X(g(y)) \cdot (-g'(y))$. Somit ergibt sich die Aussage. □

Beispiele:

1) Es sei $h(x) = ax + b$ mit $a > 0, Y = aX + b$.
   Dann ist

$$f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) \text{ und } f_X(x) = a \cdot f_Y(ax+b).$$

2) Ist $D^2X < \infty$, so bezeichnet man

$$X^* := \frac{X - EX}{\sqrt{D^2X}}$$

als die zu X gehörende standardisierte Zufallsgröße. Es gilt

$$EX^* = 0,\ D^2X^* = 1 \text{ und}$$

$$X = \sqrt{D^2X}\ X^* + EX.$$

Somit haben wir

$$f_X(x) = \frac{1}{\sqrt{D^2X}} f_{X^*}\left(\frac{x - EX}{\sqrt{D^2X}}\right).$$

3) Es sei X eine $N(\mu, \sigma^2)$-verteilte Zufallsgröße. Dann besitzt $Y = \exp(X)$ eine Dichte der Form

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}y} \exp\left[-\frac{1}{2\sigma^2}(lny - \mu)^2\right],\ y > 0$$

$$f_Y(y) = 0, \qquad\qquad y \leq 0.$$

Die Verteilung mit dieser Dichte nennt man *logarithmische Normalverteilung* mit den Parametern μ und σ^2. Es gilt

$$EY = \mathrm{e}^{\mu + \frac{\sigma^2}{2}},\ \mathrm{D^2Y} = \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2} - 1)$$

Die eindimensionale Normalverteilung $N(\mu, \sigma^2)$

Es sei $X \sim N(\mu, \sigma^2)$. Dann gilt

$$P(a < X \leq b) = P\left(\frac{a-\mu}{\sigma} < X^* \leq \frac{b-\mu}{\sigma}\right) =$$

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \quad a, b \in R_1, a < b.$$

Insbesondere erhalten wir für alle $c > 0$:

$$P(\mu - c\sigma < X \leq \mu + c\sigma) = P(|X^*| < c) =$$

$$\Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Für $c = 3$ ergibt sich:

$$P(|X - \mu| < 3\sigma) = 0,9974$$

($3 - \sigma-$Regel für die Normalverteilung)

7.4 Die Kovarianzmatrix eines zufälligen Vektors

Definition 7.28 *Es sei $X = (X_1, \ldots, X_n)^T$ ein zufälliger Vektor über $(\Omega, \mathfrak{A}, P)$ mit $E|X_i| < \infty, i = 1, \ldots, n$.*
Dann heißt der Vektor μ, definiert durch
$\mu := (\mu_1, \ldots, \mu_n)^T,\ \mu_i = EX_i, i = 1, \ldots, n,$
der Erwartungswertvektor von X. Er wird auch mit EX bezeichnet:

$$EX := (EX_1, \ldots, EX_n)^T.$$

Gilt $EX_i^2 < \infty, i = 1, \ldots, n$, so ist wegen der Cauchy-Schwarz'schen Ungleichung (7.9) auch $E|X_i X_j| < \infty, i, j = 1, \ldots, n$. Folglich sind alle Kovarianzen $Kov(X_i, X_j)$ mit $i, j = 1, \ldots, n$ endlich und es gilt:

$$Kov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) = EX_iX_j - \mu_i\mu_j.$$

Offenbar gilt $Kov(X_i, X_i) = D^2 X_i = \sigma_i^2$.

Definition 7.29 *Die Matrix* $\sum_X := (Kov(X_i, X_j))_{i,j=1,\ldots,n}$ *heißt Kovarianzmatrix des zufälligen Vektors* X*.*

Mit der Schreibweise

$$E(XX^T) := (EX_iX_j)_{i,j=1,\ldots,n}$$

gilt

$$\sum\nolimits_X = E[(X-\mu)(X-\mu)^T].$$

Auf Grund der Linearität der Erwartungswertbildung haben wir

$$\sum_X = E(XX^T) - \mu\mu^T = (E(X_iX_j) - EX_iEX_j)_{i,j=1,\ldots,n}. \tag{7.25}$$

Aussage 7.30 *Die Kovarianzmatrix* $\sum_X$ *ist symmetrisch und nichtnegativ definit. Für jeden Vektor* $a = (a_1, \ldots, a_n)^T \in R_n$ *gilt*

$$E(a^T X) = a^T(EX) \text{ und } D^2(a^T X) = E(a^T(X-\mu))^2 = a^T \sum_X a \geq 0. \tag{7.26}$$

Sind die $X_1, \ldots, X_n$ *paarweise unkorreliert, so ist* $\sum_X$ *eine Diagonalmatrix, und umgekehrt.*

Beweis:

$$E(a^T X) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i EX_i = a^T EX.$$

Die Symmetrie von $\sum_X$ folgt aus $Kov(X_i, X_j) = Kov(X_j, X_i)$. Für jedes $a \in R_n$ haben wir auf Grund der Linearität der Erwartungswertbildung

$$E(a^T X) = a^T EX \quad \text{und}$$

$$a^T \sum_X a = a^T E[(X-\mu)(X-\mu)^T]a =$$

$$E[(a^T(X-\mu))\cdot((X-\mu)^T a)] = E(a^T(X-\mu))^2 =$$

$$D^2(a^T X) \geq 0.$$

Insbesondere bedeutet dies die nichtnegative Definitheit von $\sum_X$. Der letzte Teil der Aussage ist offensichtlich. □

Die Kovarianzmatrix $\sum_X$ ist das mehrdimensionale Analogon zur Varianz σ_X^2 für reellwertige Zufallsgrößen X. Im mehrdimensionalen Fall ist die Varianz des zufälligen Vektors richtungsabhängig und i.a. nicht mehr durch eine einzige Zahl zu charakterisieren. Ist $\mathrm{e} = (\mathrm{e}_1, \ldots, \mathrm{e}_\mathrm{n})^\mathrm{T}$ ein Vektor der Länge Eins, so ist nach der vorangegangenen Aussage $\mathrm{e}^\mathrm{T} \sum_\mathrm{X} \mathrm{e} = \mathrm{E}((\mathrm{e}^\mathrm{T}(\mathrm{X}-\mu))^2)$ die Varianz der Projektion $\mathrm{e}^T X$ von X auf die durch e gegebene Richtung.

Lineare Transformationen

Die folgende Aussage wird in der linearen Algebra bewiesen.

Aussage 7.31 *Die Kovarianz $\sum_X$ ist singulär genau dann, wenn es einen Vektor $x = (x_1, \ldots, x_n)^T \neq 0$ gibt mit $x^T \sum_X x = 0$.*

Für jedes solche x gilt also wegen (7.26), dass $D^2(x^T X) = 0$ gilt.

Ist Y eine lineare Transformation des n-dimensionalen zufälligen Vektors X, d. h. gilt $Y = AX + b$ für eine $m \times n$-Matrix A und einen m-dimensionalen Vektor b, so ist

$$EY = AEX + b \text{ und } \sum\nolimits_Y = A \sum\nolimits_X A^T. \tag{7.27}$$

Da $\sum_X$ symmetrisch ist, gibt es eine orthogonale Matrix $\mathscr{O}$, so dass $\mathscr{O} \sum_X \mathscr{O}^T = D$ eine Diagonalmatrix ist:

$$D = \begin{pmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix}$$

Die Diagonalelemente d_i sind die Eigenwerte von $\sum_X$ und nichtnegativ wegen der nichtnegativen Definitheit von $\sum_X$. Der zufällige Vektor $Y := \mathscr{O}X$ besitzt gemäß (7.27) mit $\sum_Y = E\mathscr{O}XX^T\mathscr{O}^T$die Matrix D als Kovarianzmatrix. Seine Komponenten sind somit unkorreliert.

Regressionsgerade

Es sei (U, V) ein zufälliger Vektor reellwertiger Zufallsgrößen U und V.

Genau wie im Kapitel 4.5 definiert man die Regressionsgerade für V auf der Basis von U durch

$$y = EV + Kor(U, V)\frac{\sigma_2}{\sigma_1}(x - EU)$$

Die Zufallsgröße $\hat{V}$, definiert durch

$$\hat{V} := EV + Kor(U, V)\frac{\sigma_2}{\sigma_1}(U - EV)$$

ist die im quadratischen Mittel beste Vorhersage von V auf der Basis von U, d. h., es gilt

$$E(V - \hat{V})^2 = \min_{a,b \in R_1} E(V - aU - b)^2.$$

Diese Regressionsgerade ist für alle Paare reellwertiger Zufallsgrößen (U, V) definiert, für die $\sigma_1^2 = D^2U < \infty$ und $\sigma_2^2 = D^2V < \infty$ gilt.

Für den Vorhersagefehler $V - \hat{V}$ erhalten wir

$$E(V - \hat{V}) = 0 \text{ und}$$

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

Außerdem haben wir

$$Kov(U, V - \hat{V}) = E((U - EU)[(V - EV) - Kor(U, V)\frac{\sigma_1}{\sigma_2}(U - EU)) =$$

$$Kov(U, V) - \sigma_1\sigma_2 Kor(U, V) = 0.$$

7.5 Dichten mehrdimensionaler Verteilungen

In diesem Punkt studieren wir zufällige n-dimensionale Vektoren, die eine Dichte besitzen.

Definition 7.32 *Ist Q ein Wahrscheinlichkeitsmaß auf $(R_n, \mathscr{B}_n)$, und existiert eine Borelfunktion f auf R_n, so dass*
mit der Bezeichnung $(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_n]$, wobei $x = (x_1, x_2, \ldots, x_n)$ sei, gilt

$$Q((-\infty, x]) = \underset{(-\infty,x]}{\int\int \ldots \int} f(x) dx_1, dx_2 \ldots dx_n \,,\; x \in R_n,$$

dann heißt f eine Dichte des Maßes Q. Ist $Q = P^X$ für einen n-dimensionalen zufälligen Vektor X, so nennt man f auch die Dichte von X.

Dabei versteht sich das Integral als Integral bezüglich des n-dimensionalen Lebesguemaßes $\lambda(dx) = \lambda(dx_1, \ldots, dx_n) = dx_1 dx_2 \ldots dx_n$.

Auch hier haben wir für jedes $B \in \mathscr{B}_n$ die Gleichung

$$Q(B) = \int_{R_n} f(y) \mathbb{1}_B(y) dy =: \int_B f(y) dy$$

Analog zum Fall des R_1 ist eine nichtnegative Borel-messbare Funktion f auf R_n Dichte einer Wahrscheinlichkeitsverteilung auf $(R_n, \mathscr{B}_n)$ genau dann, wenn

$$\int_{R_n} f(x) dx = 1$$

gilt.

In diesem Fall bestimmt f die Verteilung eindeutig, andererseits ist die Dichte f einer n-dimensionalen Wahrscheinlichkeitsverteilung bis auf eine Menge vom n-dimensionalen Lebesguemaß Null eindeutig bestimmt.

Aussage 7.33 (Erwartungswertregel) *Es seien X ein n-dimensionaler zufälliger Vektor mit der Dichte f und h eine Borel-messbare reellwertige Funktion auf $(R_n, \mathscr{B}_n)$. Dann gilt:*

a) *$h(\cdot)$ ist bezüglich P^X integrierbar genau dann, wenn $h(\cdot)f(\cdot)$ bezüglich des n-dimensionalen Lebesguemaßes integrierbar ist,*

b) *in diesem Fall gilt*

$$Eh(X) = \int_{R_n} h(x)P^X(dx) = \int_{R_n} h(x)f(x)dx.$$

Wir haben im Fall diskreter Verteilungen die Kovarianz zweier Zufallsgrößen in Kapitel 4 berechnet. Hier wollen wir die entsprechenden Formeln für den Fall angeben, dass der zufällige Vektor X eine Dichte f besitzt.

In diesem Fall gilt nach der Erwartungswertregel

$$\mu_i = EX_i = \int_{R_n} x_i f(x_1, \ldots, x_n)dx_1, \ldots, x_n, \ i = 1, \ldots, n,$$

$$Kov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) =$$

$$\int_{R_n} (x_i - \mu_i)(x_j - \mu_j)f(x_1, \ldots, x_n)dx_1 \ldots dx_n =$$

$$\int_{R_n} x_i x_j f(x_1, \ldots, x_n)dx_1 \ldots dx_n - \mu_i\mu_j.$$

Wie man Integrale über Funktionen im R_n ausrechnet, werden wir im folgenden Kapitel 8 kennen lernen.

Wir beschränken uns im Weiteren auf den Fall $n = 2$.

Es sei $X = (Y, Z)^T$ ein zufälliger Vektor mit Werten in $(R_2, \mathscr{B}_2)$ und der Dichte f.

Aussage 7.34

a) *Y und Z haben Dichten f_Y bzw. f_Z, die sich mittels f wie folgt berechnen lassen:*

$$f_Y(y) = \int\limits_{R_1} f(y,z)dz, \quad y \in R_1$$

$$f_Z(z) = \int\limits_{R_1} f(y,z)dy, \quad z \in R_1$$

b) Y und Z sind genau dann voneinander unabhängig, falls

$$f(y,z) = f_Y(y)f_Z(z) \quad (y,z) \in R_2, \quad \lambda_2 - \textit{fast überall}$$

c) für jedes y mit $f_Y(y) > 0$ ist durch

$$f_{Y=y}(z) := \frac{f(y,z)}{f_Y(y)} \quad , \quad z \in R_1$$

eine Dichte definiert. Sie heißt "bedingte Dichte von Z unter der Bedingung $Y = y$."

Man nennt f_Y und f_Z die *Randverteilungsdichten* von f.

Beweis:

a) $P(Y \le y) = P(Y \le y, Z \in R_1) =$

$$\iint\limits_{(-\infty,y]\times R_1} f(s,t)dsdt = \int\limits_{(-\infty,y]} \left(\int\limits_{R_1} f(s,t)dt \right) ds.$$

Folglich gilt die erste Formel von a), analog folgt die zweite (es wurde der Satz von Fubini benutzt.)

b) Wenn $f = f_Y f_Z$, so ist

$$P(Y \in B, Z \in C) = \iint\limits_{B\times C} f(y,z)dydz =$$

$$\int\limits_{R_1} \mathbb{1}_C f_Z(z)\Big(\int\limits_{R_1} \mathbb{1}_B(y) f_Y(y)dy\Big)dz =$$

$$P(Y \in B)P(Z \in C), \quad B, C \in \mathscr{B}_1.$$

Also sind Y und Z unabhängig.

Umgekehrt, sind Y und Z unabhängig, so gilt

$$F(y,z) = P(Y \leq y, Z \leq z) = F_Y(y)F_Z(z) =$$

$$\int\limits_{(-\infty,y]} f_Y(y)dy \int\limits_{(-\infty,z]} f_Z(z)dz = \iint\limits_{(-\infty,y]\times(-\infty,z]} f_Y(y)f_Z(z)dydz$$

(Fubini; Tonelli, Hobson)). Wegen der Eindeutigkeit der Dichte besitzt $(Y,Z)^T$ eine Dichte f, und es gilt

$$f(y,z) = f_Y(y) \cdot f_Z(z) \quad , \quad \lambda - \text{f.ü.}$$

c) Es gilt $f_{Y=y}(z) \geq 0$ und $\int\limits_{R_1} f_{Y=y}(z)dz = 1$.

Bemerkung: Interpretation von c):

Es sei $f(y,z)$ stetig und streng positiv in (y_0, z_0).

Dann ist $f(x,y) > 0$ in Umgebung U von (y_0, z_0) (z.B. $U = (y_0 - \triangle, y_0 + \triangle) \times (z_0 - \triangle, z_0 + \triangle)$ für genügend kleines $\triangle > 0$).

Wir erhalten für jedes $z \in R_1$

$$P(Z \leq z | Y \in (y_0 - \triangle, y_0 + \triangle)) = \frac{P(Y \in (y_0 - \triangle, y_0 + \triangle), Z \leq z)}{P(Y \in (y_0 - \triangle, y_0 + \triangle))} =$$

$$\int\limits_{(-\infty,z]} \Big(\int\limits_{(y_0-\triangle,y_0+\triangle)} f(s,t)ds \Big) dt \Big/ \int\limits_{R_1} \Big(\int\limits_{(y_0-\triangle,y_0+\triangle)} f(s,t)ds \Big) dt =$$

$$\sim \int\limits_{(-\infty,z]} (2\triangle f(y_0,t)dt \Big/ \int\limits_{R_1} 2\triangle f(y_0,t)dt$$

$$= \int\limits_{(-\infty,z]} f_{Y=y_0}(s)ds$$

und sehen darin eine Interpretation von $f_{Y=y_0}(z)$ als Dichte von Z unter der Bedingung $Y = y_0$.

Beispiel 7.35 $X = (X_1, X_2)^T$ besitze eine 2-dimensionale Normalverteilung mit den Parametern $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \varrho$. Dann gilt

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(\frac{-1}{2(1-\varrho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\varrho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] \right.$$

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2 \right)$$

$$EX_i = \int\limits_{R_2} x_i f_X(x_1,x_2)dx_1dx_2 = \int\limits_{R_1} x_i f_{X_i}(x_i)dx_i = \mu_i$$

$$D^2X_i = \int\limits_{R_2} (x_i-\mu_i)^2 f_X(x_1,x_2)dx_1dx_2 = \int\limits_{R_1} (x_i-\mu_i)^2 f_{X_i}(x_i)dx_i = \sigma_i^2$$

$$Kov(X_1,X_2) = \int\limits_{R_2} (x_i-\mu_1)(x_2-\mu_2) f_X(x_1,x_2)dx_1dx_2 =$$

$$\int\limits_{R_1} (x_2-\mu_2) \Big(\int\limits_{R_1} (x_1-\mu_1) \cdot f_X(x_1,x_2)dx_1 \Big) dx_2 = \varrho\sigma_1\sigma_2.$$

Damit ist die Bedeutung der Parameter der 2-dimensionalen Normalverteilung geklärt. Folglich haben wir für die Kovarianzmatrix des Vektors X

$$\sum\nolimits_X = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

und ϱ ist gleich dem Korrelationskoeffizienten $Kor(X_1, X_2)$.

Man prüft leicht nach, dass sich die Dichte f_X in diesem Beispiel folgendermaßen schreiben läßt ($\mu = (\mu_1, \mu_2)^T, x = (x_1, x_2)^T$):

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left[-\frac{1}{2}(x-\mu)^T \sum\nolimits_X^{-1}(x-\mu)\right]$$

Für die bedingte Dichte $f_{X_1=x_1}(x_2)$ ergibt sich

$$f_{X_1=x_1}(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^*}} \exp\left[-\frac{1}{2\sigma_2^*}(x_2-\mu_2^*)^2\right]$$

mit

$$\mu_2^* = \mu_2 + \varrho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1) \text{ und}$$

$$\sigma_2^{*2} = \sigma_2^2(1-\varrho^2)$$

Beachte, dass σ_2^{*2} nicht von x_1 abhängt.

Die Komponenten X_1 und X_2 sind genau dann unabhängig, falls sie unkorreliert sind, also $\varrho = 0$ gilt. In der Tat, genau in diesem Fall gilt

$$f_X(x) = f_{X_1}(x_1)f_{X_2}(x_2), \quad x = (x_1, x_2)^T.$$

Die Transformationsformel für n-dimensionale Dichten (Aussage (3.59) und Beispiel (3.60) bleiben in diesem allgmeinen Fall gültig.

Index