Inhaltsverzeichnis

Kapitel VI

Erwartungswert und Varianz

Erwartungswert und Varianz sind fundamentale Kenngrößen der Verteilung einer reellwertigen Zufallsvariable.

1 Der Erwartungswert

Erwartungswert: „mittlerer Wert“ einer Zufallsvariablen, „Schwerpunkt“ ihrer Verteilung. Die allgemeine Definition basiert auf dem abstraktem Lebesgue-Integral, siehe Georgii (2007, Abschn. 4.1.2).

Bemerkung 1. Die Menge der Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ bildet einen Vektorraum, der auch abgeschlossen unter Multiplikation ist. Auf dem Untervektorraum $\mathfrak{L}_1 = \mathfrak{L}_1(\Omega, \mathfrak{A}, P)$ der *integrierbaren Zufallsvariablen* erklärt man das Integral

$$\mathrm{E}(X) = \int_\Omega X(\omega)\, dP(\omega),$$

genannt *Erwartungswert* von X. Für $A \in \mathfrak{A}$ gilt $1_A \in \mathfrak{L}_1$ und

$$\mathrm{E}(1_A) = P(A).$$

Die Abbildung $\mathrm{E} : \mathfrak{L}_1 \to \mathbb{R}$ ist linear und monoton, d.h. für $X, Y \in \mathfrak{L}_1$ und $\alpha, \beta \in \mathbb{R}$ gilt

- $\mathrm{E}(\alpha X + \beta Y) = \alpha\,\mathrm{E}(X) + \beta\,\mathrm{E}(Y)$,
- $X \le Y \Rightarrow \mathrm{E}(X) \le \mathrm{E}(Y)$.

Ferner gilt für Zufallsvariablen X und Y auf $(\Omega, \mathfrak{A}, P)$

- $|X| \le Y \wedge Y \in \mathfrak{L}_1 \Rightarrow X \in \mathfrak{L}_1$,

- falls $X \geq 0$: $X \in \mathfrak{L}_1 \wedge \mathrm{E}(X) = 0 \Leftrightarrow P(\{X > 0\}) = 0$.

Wir betrachten die Spezialfälle

(i) X diskret, also $P(\{X \in D\}) = 1$ für eine abzählbare Menge, $D \subset \mathbb{R}$

(ii) X absolutstetig mit Dichte f_X.

Die folgenden Sätze dienen uns als Definitionen, siehe Irle (2001, Kap. 8) oder Georgii (2007, Abschn. 4.1).

Satz 2. Im Fall (i) gilt $X \in \mathfrak{L}_1$ genau dann, wenn $\sum_{x \in D} |x| \cdot P(\{X = x\}) < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X) = \sum_{x \in D} x \cdot P(\{X = x\}).$$

Satz 3. Im Fall (ii) gilt $X \in \mathfrak{L}_1$ genau dann, wenn $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

Bemerkung 4. Die Integrierbarkeit von X und gegebenenfalls $\mathrm{E}(X)$ hängt nur von der Verteilung P_X ab.

Satz 5.

$$\begin{aligned}
X \sim \mathbf{B}(n,p) \quad &\Rightarrow \quad \mathrm{E}(X) = n \cdot p \\
X \sim \mathbf{G}(p) \quad &\Rightarrow \quad \mathrm{E}(X) = 1/p \\
X \sim \mathbf{P}(\lambda) \quad &\Rightarrow \quad \mathrm{E}(X) = \lambda \\
X \sim \mathbf{U}([a,b]) \quad &\Rightarrow \quad \mathrm{E}(X) = (a+b)/2 \\
X \sim \mathbf{Exp}(\lambda) \quad &\Rightarrow \quad \mathrm{E}(X) = 1/\lambda \\
X \sim \mathbf{N}(\mu, \sigma^2) \quad &\Rightarrow \quad \mathrm{E}(X) = \mu
\end{aligned}$$

Beweis. Für $X \sim \mathbf{B}(1,p)$ gilt

$$\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

Für $X \sim \mathbf{B}(n,p)$ könne wir wegen Bemerkung 4 und Satz III.4.8 oBdA annehmen, daß

$$X = \sum_{i=1}^{n} X_i$$

mit $X_1, \dots, X_n$ iid und $X_1 \sim \mathbf{B}(1,p)$. Also

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = n \cdot \mathrm{E}(X_1) = n \cdot p.$$

Beachte, daß hier die Unabhängigkeit nicht verwendet wurde.
Sei f_X die Dichte von $X \sim \mathbf{Exp}(\lambda)$. Dann gilt

$$\int_{-\infty}^{\infty} |x| \cdot f_X(x)\, dx = \int_0^{\infty} x \cdot \lambda \cdot \exp(-\lambda x)\, dx = -x \cdot \exp(-\lambda x)\Big|_0^{\infty} + \int_0^{\infty} \exp(-\lambda x)\, dx$$
$$= \frac{-1}{\lambda} \exp(-\lambda x)\Big|_0^{\infty} = \frac{1}{\lambda}.$$

Für die restlichen Verteilungen $\boxed{\text{Übung}}$. ☐

Beispiel 6. Betrachte Rückkehrzeit

$$\tau_\infty := \inf\{t \in \mathbb{N} : S_t = 0\}$$

der symmetrischen Bernoulli-Irrfahrt $(S_t)_{t \in \mathbb{N}_0}$ mit unendlichem Zeithorizont. Es gilt $P(\{\tau_\infty < \infty\}) = 1$, siehe Satz III.5.25, aber $\tau_\infty \notin \mathfrak{L}_1$, siehe $\boxed{\text{Übung}}$.

Nun: Hilfsmittel zur Berechnung von Erwartungswerten.

Bezeichnung. $(\Omega_j)_{j \in J}$ *abzählbare Partition* von Ω, falls

(i) J abzählbar,

(ii) $\Omega_j \in \mathfrak{A}$ für alle $j \in J$,

(iii) $\Omega_j \cap \Omega_\ell = \emptyset$ für $j \neq \ell$,

(iv) $\Omega = \bigcup_{j \in J} \Omega_j$.

Lemma 7. Sei X eine Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$, und sei $(\Omega_j)_{j \in J}$ eine abzählbare Partition von Ω. Gelte

$$\forall j \in J\ \exists c_j \in \mathbb{R}\ \forall \omega \in \Omega_j :\ X(\omega) = c_j.$$

Dann gilt $X \in \mathfrak{L}_1$ genau dann, wenn

$$\sum_{j \in J} |c_j| \cdot P(\Omega_j) < \infty.$$

Gegebenenfalls folgt

$$\mathrm{E}(X) = \sum_{j \in J} c_j \cdot P(\Omega_j).$$

Beweis. Für
$$D := \{c_j : j \in J\}$$
gilt
$$P(\{X \in D\}) = 1.$$
Setze $J_x = \{j \in J : c_j = x\}$ für $x \in D$. Dann
$$\begin{aligned}\sum_{x \in D} |x| \cdot P(\{X = x\}) &= \sum_{x \in D} |x| \cdot \sum_{j \in J_x} P(\Omega_j) = \sum_{x \in D} \sum_{j \in J_x} |c_j| \cdot P(\Omega_j) \\ &= \sum_{j \in J} |c_j| \cdot P(\Omega_j).\end{aligned}$$
Berechnung des Erwartungswertes analog ohne Beträge. □

Bemerkung 8. Satz 2 beruht auf der abzählbaren Partition $(\{X = x\})_{x \in D}$.

Korollar 9. Falls $(\Omega, \mathfrak{A}, P)$ diskret, gilt $X \in \mathfrak{L}_1$ genau dann, wenn
$$\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) < \infty.$$
Gegebenenfalls folgt
$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}).$$

Beweis. Wähle $J := \Omega$, $\Omega_j := \{j\}$ und $c_j := X(j)$ in Lemma 7. □

Definition 10. $g : \mathbb{R}^d \to \mathbb{R}$ heißt *Borel-meßbar*, falls
$$\forall A \in \mathfrak{B}_1 : \; g^{-1}(A) \in \mathfrak{B}_d.$$

Bemerkung 11.

(i) Die Menge der Borel-meßbaren Funktionen $\mathbb{R}^d \to \mathbb{R}$ bildet einen Vektorraum, der auch abgeschlossen unter Multiplikation ist.

(ii) Jede stetige Funktion $\mathbb{R}^d \to \mathbb{R}$ ist Borel-meßbar.

(iii) $g = 1_A$ mit $A \in \mathfrak{B}_d$ ist Borel-meßbar.

Siehe Irle (2001, Kap. 7).

Nun: ein Transformationssatz für den Erwartungswert.

Lemma 12. Sei X ein d-dimensionaler Zufallsvektor mit Dichte f_X und $h : \mathbb{R}^d \to \mathbb{R}$ Borel-meßbar. Genau dann gilt $h(X) \in \mathfrak{L}_1$, wenn $\int_{\mathbb{R}^d} |h(x)| \cdot f_X(x)\, dx < \infty$. Gegebenenfalls folgt
$$\mathrm{E}(h(X)) = \int_{\mathbb{R}^d} h(x) \cdot f_X(x)\, dx < \infty.$$

Beweis. Siehe Irle (2001, Satz 8.25). □

Beachte: Lemma 12 gilt ohne die Annahme, daß $h(X)$ absolutstetig verteilt ist.

Satz 13. Sind $X, Y \in \mathfrak{L}_1$ unabhängig, so folgt $X \cdot Y \in \mathfrak{L}_1$ und

$$\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Beweis. Siehe Irle (2001, Satz 10.16). Hier Beweis unter der zusätzlichen Annahme, daß X, Y diskrete Zufallsvariablen sind.
Wähle $D \subset \mathbb{R}$ abzählbar mit $P(\{X \in D\}) = P(\{Y \in D\}) = 1$, setze

$$\begin{aligned} \Omega_{(x,y)} &= \{(X,Y) = (x,y)\}, \qquad && (x,y) \in D^2, \\ \Omega_* &= \{X \notin D\} \cup \{Y \notin D\} \end{aligned}$$

sowie

$$X' = 1_D \cdot X \quad \text{und} \quad Y' = 1_D \cdot Y.$$

Dann

$$P(\{X \cdot Y \neq X' \cdot Y'\}) \leq P(\{X \notin D\} \cup \{Y \notin D\}) = 0.$$

Also folgt $P_{X \cdot Y} = P_{X' \cdot Y'}$ und somit $X \cdot Y \in \mathfrak{L}_1 \Leftrightarrow X' \cdot Y' \in \mathfrak{L}_1$.
Ferner gilt

$$\sum_{(x,y) \in D^2} |x \cdot y| \cdot P(\Omega_{(x,y)}) + 0 \cdot P(\Omega_*) = \sum_{x \in D} |x| \cdot P(\{X = x\}) \cdot \sum_{y \in D} |y| \cdot P(\{X = y\}) < \infty,$$

und Lemma 7 zeigt $X' \cdot Y' \in \mathfrak{L}_1$.
Berechnung des Erwartungswertes analog ohne Beträge.
Im Spezialfall absolutstetiger Zufallsvariablen X, Y verwendet man den Satz von Fubini und Satz V.2.21.(i). □

2 Varianz und Kovarianz

Varianz: „Streuungsmaß" für die Verteilung einer Zufallsvariablen.

Definition 1. X *quadratisch integrierbar*, falls $X^2 \in \mathfrak{L}_1$. Bez.: $\mathfrak{L}_2 = \mathfrak{L}_2(\Omega, \mathfrak{A}, P)$ Menge der quadratisch integrierbaren Zufallsvariablen.

Satz 2. $\mathfrak{L}_2$ ist Untervektorraum von $\mathfrak{L}_1$.

Beweis. Verwende $|X| \leq 1 + X^2$ und $(X+Y)^2 \leq 2\,X^2 + 2\,Y^2$. □

Bemerkung 3. Die quadratische Integrierbarkeit von X und gegebenenfalls $\mathrm{E}(X^2)$ hängt nur von der Verteilung P_X ab.

Satz 4. Im Fall (i) gilt $X \in \mathfrak{L}_2$ genau dann, wenn $\sum_{x\in D} x^2 \cdot P(\{X = x\}) < \infty$. Gegebenenfalls folgt
$$\mathrm{E}(X^2) = \sum_{x\in D} x^2 \cdot P(\{X = x\}).$$

Beweis. Wende Lemma 1.7 mit $J := D$, $\Omega_j := \{X = j\}$ und $c_j := j^2$ an. □

Satz 5. Im Fall (ii) gilt $X \in \mathfrak{L}_2$ genau dann, wenn $\int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx < \infty$. Gegebenenfalls folgt
$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx.$$

Beweis. Wende Lemma 1.12 mit $h(x) := x^2$ an. □

Definition 6. Für $X \in \mathfrak{L}_2$ heißt
$$\mathrm{Var}(X) := \mathrm{E}(X - \mathrm{E}(X))^2$$
die *Varianz* und $\sqrt{\mathrm{Var}(X)}$ die *Standardabweichung* von X.

Nun: Abschätzung für die Konzentration einer Zufallsvariable um ihren Erwartungswert.

Bemerkung 7. Für $X \in \mathfrak{L}_2$ gilt
$$\mathrm{Var}(X) = 0 \Leftrightarrow P(\{X = \mathrm{E}(X)\}) = 1.$$

Satz 8 (Tschebyschev-Ungleichung)**.** Für $X \in \mathfrak{L}_2$ und $\varepsilon > 0$ gilt
$$P(\{|X - \mathrm{E}(X)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(X)$$

Beweis. Für $A := \{|X - \mathrm{E}(X)| \geq \varepsilon\} \in \mathfrak{A}$ gilt
$$\varepsilon^2 \cdot 1_A \leq (X - \mathrm{E}(X))^2 \cdot 1_A \leq (X - \mathrm{E}(X))^2.$$
Es folgt
$$\varepsilon^2 \cdot P(A) = \varepsilon^2 \cdot \mathrm{E}(1_A) \leq \mathrm{E}(X - \mathrm{E}(X))^2.$$
□

Satz 9. Für $X \in \mathfrak{L}_2$ und $\alpha, \beta \in \mathbb{R}$ gilt

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$,

(ii) $\mathrm{Var}(\alpha \cdot X + \beta) = \alpha^2 \cdot \mathrm{Var}(X)$.

Beweis. Ad (i): Es gilt

$$(X - \mathrm{E}(X))^2 = X^2 - 2 \cdot X \cdot \mathrm{E}(X) + (\mathrm{E}(X))^2.$$

Es folgt

$$\mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - 2 \cdot (\mathrm{E}(X))^2 + (\mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$$

Ad (ii): Es gilt

$$\alpha \cdot X + \beta - \mathrm{E}(\alpha \cdot X + \beta) = \alpha \cdot (X - \mathrm{E}(X)).$$

Es folgt

$$\mathrm{Var}(\alpha \cdot X + \beta) = \mathrm{E}(\alpha^2 \cdot (X - \mathrm{E}(X))^2) = \alpha^2 \cdot \mathrm{Var}(X).$$

□

Bemerkung 10. Für $X, Y \in \mathfrak{L}_2$ gilt $X \cdot Y \in \mathfrak{L}_1$. Zum Beweis verwende man $|X \cdot Y| \leq X^2 + Y^2$.

Definition 11. Betrachte $X, Y \in \mathfrak{L}_2$.

(i) Die *Kovarianz* von X und Y ist definiert durch

$$\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}(X)) \cdot (Y - \mathrm{E}(Y))).$$

(ii) X, Y heißen *unkorreliert*, falls $\mathrm{Cov}(X, Y) = 0$.

(iii) Falls $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$, so ist der *Korrelationskoeffizient* von X und Y definiert durch

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Bemerkung 12.

(i) Für $X, Y \in \mathfrak{L}_2$ gilt

$$\mathrm{Cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

(ii) $X, Y \in \mathfrak{L}_2$ unabhängig $\Rightarrow X, Y$ unkorreliert, siehe Satz 1.13. Die Umkehrung ist falsch, siehe $\boxed{\text{Übung}}$.

Satz 13 (Formel von Bienaymé)**.** Falls $X_1, \ldots, X_n \in \mathfrak{L}_2$ paarweise unkorreliert,

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Beweis. Setze $Y_i := X_i - \mathrm{E}(X_i)$ („zentrieren“). Für $i \neq j$ gilt $\mathrm{E}(Y_i \cdot Y_j) = 0$. Also

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) &= \mathrm{Var}\left(\sum_{i=1}^n Y_i\right) = \mathrm{E}\left(\sum_{i=1}^n Y_i\right)^2 \\ &= \sum_{i,j=1}^n \mathrm{E}(Y_i \cdot Y_j) = \sum_{i=1}^n \mathrm{E}(Y_i^2) = \sum_{i=1}^n \mathrm{Var}(X_i).\end{aligned}$$

□

Beispiel 14. Für $X_1 \sim \mathbf{B}(1, 1/2)$ und $X_2 = -X_1$ gilt $\mathrm{Var}(X_1 + X_2) = 0$ und $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1/4$.

Satz 15.

$$\begin{aligned}X \sim \mathbf{B}(n,p) \quad &\Rightarrow \quad \mathrm{Var}(X) = n \cdot p \cdot (1-p) \\ X \sim \mathbf{G}(p) \quad &\Rightarrow \quad \mathrm{Var}(X) = (1-p)/p^2 \\ X \sim \mathbf{P}(\lambda) \quad &\Rightarrow \quad \mathrm{Var}(X) = \lambda \\ X \sim \mathbf{U}([a,b]) \quad &\Rightarrow \quad \mathrm{Var}(X) = (b-a)^2/12 \\ X \sim \mathbf{Exp}(\lambda) \quad &\Rightarrow \quad \mathrm{Var}(X) = 1/\lambda^2 \\ X \sim \mathbf{N}(\mu, \sigma^2) \quad &\Rightarrow \quad \mathrm{Var}(X) = \sigma^2\end{aligned}$$

Beweis. Für $X \sim \mathbf{B}(1,p)$ gilt

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = p - p^2 = p \cdot (1-p).$$

Für $X \sim \mathbf{B}(n,p)$ können wir wegen der Bemerkungen 1.4 und 3 und Satz III.4.8 oBdA annehmen, daß

$$X = \sum_{i=1}^n X_i$$

mit $X_1, \dots, X_n$ iid und $X_1 \sim \mathbf{B}(1,p)$. Mit Satz 13 folgt

$$\mathrm{Var}(X) = n \cdot p \cdot (1-p).$$

Für $X \sim \mathbf{Exp}(\lambda)$ gilt

$$\begin{aligned}&\int_0^\infty x^2 \cdot \lambda \cdot \exp(-\lambda x)\,dx \\ &\quad = -x^2 \cdot \exp(-\lambda x)\Big|_0^\infty + \int_0^\infty 2x \cdot \exp(-\lambda x)\,dx \\ &\quad = \frac{2}{\lambda} \cdot \mathrm{E}(X) = \frac{2}{\lambda^2}.\end{aligned}$$

Demnach gilt $\mathrm{E}(X^2) = \dfrac{2}{\lambda^2}$ und

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Für die restlichen Verteilungen $\boxed{\text{Übung}}$. □

Satz 16 (Cauchy-Schwarzsche-Ungleichung)**.** Für $X, Y \in \mathfrak{L}_2$ gilt

$$|\,\mathrm{E}(X \cdot Y)| \le \sqrt{\mathrm{E}(X^2) \cdot \mathrm{E}(Y^2)}.$$

Beweis. $\boxed{\text{Übung.}}$ □

Bemerkung 17. Für $X, Y \in \mathfrak{L}_2$ mit $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$ gilt

$$-1 \le \rho(X, Y) \le 1,$$

Satz 18. Für $X, Y \in \mathfrak{L}_2$ mit $\mathrm{Var}(X) > 0$ seien

$$b^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \qquad a^* = \mathrm{E}(Y) - b^* \cdot \mathrm{E}(X).$$

Dann

$$\mathrm{E}(Y - (a^* + b^* \cdot X))^2 = \mathrm{Var}(Y) \cdot (1 - \rho^2(X, Y)) \tag{1}$$

und

$$\mathrm{E}(Y - (a^* + b^* \cdot X))^2 \le \mathrm{E}(Y - (a + b \cdot X))^2 \tag{2}$$

für alle $a, b \in \mathbb{R}$.

Beweis. $\boxed{\text{Übung}}$. □

Bemerkung 19.

(i) Interpretation von (2): $a^* + b^* \cdot X$ ist die beste lineare Vorhersage von Y bei Beobachtung von X bzgl. des $\mathfrak{L}_2$-Abstandes.

(ii) Interpretation von (1): $\rho^2(X, Y)$ und $\mathrm{sgn}(\rho(X, Y))$ geben den Grad und die Richtung des linearen Zusammenhang von X und Y an. Extremfälle:

$$|\rho(X, Y)| = 1 \Leftrightarrow P(\{Y = a^* + b^* \cdot X\}) = 1,$$

und $\rho(X, Y) = 0$ gilt genau dann, wenn X nicht in die beste lineare Vorhersage eingeht.

(iii) Interpretation von Satz 18: orthogonale Projektion von Y auf den von 1 und X erzeugten Unterraum des Hilbertraumes L_2.

Definitionen und Bezeichnungen