3.3 Varianz und Kovarianz

Definition 3.12. X reelle ZV, $p > 0$. X besitzt p-tes Moment (auch geschrieben $X \in \mathcal{L}^p$), wenn $\mathbb{E}\big[|X|^p\big] < \infty$. $\mathbb{E}\big[X^p\big]$ heißt das p-te Moment von X.

(bzw. $\mathcal{L}^p(P)$, wenn das zugehörige W'maß P nicht aus dem Kontext klar)

Bemerkung 3.13. Für $p \geq p' \geq 1$ ist $\mathcal{L}^p \subset \mathcal{L}^{p'}$ (denn $|x|^{p'} \leq 1 + |x|^p$).

Definition 3.14. Für $X, Y \in \mathcal{L}^2$ heißt

1. $\mathrm{Var}[X] := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}[X]\big)^2$ die *Varianz* von X (manchmal schreibt man auch $\sigma_X^2 := \mathrm{Var}[X]$),

   $\sqrt{\mathrm{Var}[X]}$ die *Standardabweichung* (oder *Streuung*) von X (manchmal auch $\sigma_X = \sqrt{\sigma_X^2}$ geschrieben),

   (ein „populäres Maß" für die Variabilität von X)

2. $\mathrm{Cov}[X, Y] := \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ die *Kovarianz* von X und Y.

X und Y heißen *unkorreliert*, wenn $\mathrm{Cov}[X, Y] = 0$.

$$\mathrm{Var}[X] = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big], \qquad \mathrm{Cov}[X,Y] := \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$$

Beobachtung 3.15. 1. Wegen $|XY| \le X^2 + Y^2$ ist die Kovarianz wohldefiniert. Es gilt (offensichtlich)

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X].$$

2. Es ist

$$\begin{aligned}\mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] &= \mathbb{E}\big[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]\big]\\ &= \mathbb{E}[YX] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

(und analog für $\mathrm{Var}[X] = \mathrm{Cov}[X,X]$). $\leftarrow \mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

3. $\mathrm{Var}[X] = 0 \iff P\big(X = \mathbb{E}[X]\big) = 1$
(„$\Leftarrow$" ist klar, für „$\Rightarrow$" wende Satz 3.4, 3. an auf die ZV $(X - \mathbb{E}[X])^2$) (d.h. $\mathbb{P}((X-\mathbb{E}[X])^2=0)=1$)

4. $\mathrm{Var}[X]$ ist eine Eigenschaft der Verteilung von X, $\mathrm{Cov}[X,Y]$ ist eine Eigenschaft der gemeinsamen Verteilung von X und Y.

Beispiel 3.16. 1. $X \sim \mathrm{Ber}_p$, $\mathrm{Var}[X] = \underbrace{\mathbb{E}[X^2]}_{= p \cdot 1^2 + (1-p)\cdot 0^2 = p} - (\mathbb{E}(X))^2 = p - p^2 = p(1-p)$

2. $X \sim \mathrm{Poi}_\alpha$, so ist $\mathrm{Var}[X] = \alpha$

$$\mathbb{E}[\underbrace{X(X-1)}_{=X^2-X}] = \sum_{k=\cancel{0}\,2}^{\infty} k(k-1)e^{-\alpha}\frac{\alpha^k}{k!} = \cancel{e^{-\alpha}}\,\alpha^2 \underbrace{\sum_{k=2}^{\infty}\frac{\alpha^{k-2}}{(k-2)!}}_{=\sum_{j=0}^{\infty}\frac{\alpha^j}{j!} = \cancel{e^{\alpha}}}$$

$$= \alpha^2$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\underbrace{\mathbb{E}[X]}_{=\alpha})^2$$

$$= \underbrace{\mathbb{E}[X(X-1)]}_{=\alpha^2} + \underbrace{\mathbb{E}[X]}_{=\alpha} - \alpha^2 = \alpha$$

3. $X \sim \mathrm{Bin}_{n,p}$,

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k}$$

$$= n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{(n-2)-(k-2)} = n(n-1)p^2$$

also

$$\mathrm{Var}[X] = \mathbb{E}\big[X(X-1)\big] + \mathbb{E}[X] - \big(\mathbb{E}[X]\big)^2 = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p)$$

4. $X \sim \mathrm{Geom}_p$, $p \in [0,1]$ (d.h. $P(X = k) = p(1-p)^k$, $k \in \mathbb{N}_0$, vgl. Bsp. 1.21), so ist $\mathrm{Var}[X] = \dfrac{1-p}{p^2}$

$$\mathbb{E}[X(X-1)] = \sum_{n=\cancel{0}\,2}^{\infty} n(n-1)p(1-p)^n = (1-p)^2 p \sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2}$$

$$f(t) = \sum_{n=0}^{\infty} t^n, \quad |t| < 1 \qquad = \frac{2(1-p)^2}{p^2}$$

$$= \frac{1}{1-t}$$

$$f'(t) = \frac{1}{(1-t)^2}, \quad f''(t) = \frac{2}{(1-t)^3} = \sum_{n=2}^{\infty} n(n-1)t^{n-2}$$

$$\mathrm{Var}[X] = \mathbb{E}[X(X-1)] + \underbrace{\mathbb{E}[X]}_{=\frac{1-p}{p}} - (\mathbb{E}[X])^2$$

$$= \ldots = \frac{1-p}{p^2}$$

5. $X \sim \mathcal{N}_{\mu,\sigma^2}$, $\mathrm{Var}[X] = \sigma^2$

(es gilt $\mu = \mathbb{E}[X]$)

$$\mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2]$$

$$= \int_{\mathbb{R}} \underbrace{(x-\mu)^2}_{=(\sigma z)^2} \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{e^{-z^2/2}} dx$$

Setze $z = \frac{x-\mu}{\sigma}$, $dz = \frac{1}{\sigma} dx$ $\Updownarrow$ $x = \sigma z + \mu$

$$= \sigma^2 \int_{\mathbb{R}} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= \sigma^2 \frac{1}{\sqrt{2\pi}} \underbrace{\int_{\mathbb{R}} z \cdot \underbrace{\left(z e^{-\frac{z^2}{2}}\right)}_{= -\frac{d}{dz} e^{-z^2/2}} dz}_{= -z \cdot e^{-z^2/2} \Big|_{-\infty}^{\infty} - \underbrace{\int_{\mathbb{R}} 1 \cdot (-e^{-z^2/2}) dz}_{= \sqrt{2\pi}}} = \sigma^2$$

Satz 3.17 (Rechenregeln für Varianz und Kovarianz)**.** *Seien* $X, Y, X_1, X_2, \ldots, X_n \in \mathcal{L}^2$, $a, b, c, d \in \mathbb{R}$.

1. $aX + b, cY + d \in \mathcal{L}^2$ *und*

$$\mathrm{Cov}[aX + b, cY + d] = ac\mathrm{Cov}[X, Y],$$

insbesondere

$$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$$

(die Kovarianz ist eine Bilinearform, die Varianz ein quadratisches Funktional).

2. $\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{\substack{1\le i,j\le n\\ i\neq j}} \mathrm{Cov}[X_i, X_j]$,

insbesondere gilt für paarweise unkorrelierte $X_1, \ldots, X_n$ *also* $\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$.

3. *Sind* X *und* Y *unabhängig, so gilt* $\mathrm{Cov}[X, Y] = 0$.

4. *Es gilt*

$$\left|\mathrm{Cov}[X, Y]\right| \le \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]} \quad \textit{(Cauchy-Schwarz-Ungleichung[2])}$$

[2] nach Augustin-Louis Cauchy (1789–1857) und Hermann Amandus Schwarz (1843–1921)

Beweis. 1. Z.z: $\underbrace{\mathrm{Cov}[aX+b, cY+d]}$ $= ac\mathrm{Cov}[X, Y]$

$$\left(aX + \cancel{b} - \underbrace{\mathbb{E}[aX+b]}_{\mathbb{E}[aX] + \cancel{b}} \right)$$

$$= \mathrm{Cov}[aX, cY]$$

$$= \mathbb{E}[(aX)(cY)] - \mathbb{E}[aX] \cdot \mathbb{E}[cY]$$

$$= ac \left(\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] \right)$$

✓

2. Z.z.: $\operatorname{Var}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \operatorname{Var}[X_i] + \sum_{i \neq j}^{n} \operatorname{Cov}[X_i, X_j]$ (Induktiv mit 1. oder:)

Sei o.E. $\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n] = 0$ (sonst betr. $\tilde{X}_i = X_i - \mathbb{E}[X_i]$)

dann ist

$$\operatorname{Var}\Big[\sum_{i=1}^{n} X_i\Big] = \mathbb{E}\Big[\underbrace{\Big(\sum_{i=1}^{n} X_i\Big)^2}_{= \sum_{i,j=1}^{n} X_i \cdot X_j = \sum_{i=1}^{n} X_i^2 + \sum_{i \neq j}^{n} X_i X_j}\Big] \quad (-\,0^2)$$

$$= \sum_{i=1}^{n} \underbrace{\mathbb{E}[X_i^2]}_{\operatorname{Var}[X_i]} + \sum_{i \neq j}^{n} \underbrace{\mathbb{E}[X_i X_j]}_{= \operatorname{Cov}[X_i, X_j]}$$

3. X und Y unabhängig $\Longrightarrow \operatorname{Cov}[X, Y] = 0$

denn dann gilt $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ ✓

4. Z.z. $\left|\mathrm{Cov}[X,Y]\right| \leq \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}$ (Cauchy-Schwarz-Ungleichung)

- Falls $\mathrm{Var}[Y] = 0$, so ist $\mathbb{P}(Y - \mathbb{E}[Y] = 0) = 1$,
also auch $\mathrm{Cov}[X,Y] = \mathbb{E}\big[(X - \mathbb{E}[X])\underbrace{(Y - \mathbb{E}[Y])}_{=0}\big] = 0$
(Formel gilt als $0 = 0$)

- Falls $\mathrm{Var}[Y] > 0$, setze $\alpha := -\dfrac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]}$

$$0 \leq \underbrace{\mathrm{ar}[X + \alpha Y]}_{\mathrm{Cov}[X+\alpha Y,\, X+\alpha Y]} \cdot \mathrm{Var}[Y]$$

$$\overset{(1.)}{=} \left(\mathrm{Var}[X] + 2\alpha\,\mathrm{Cov}[X,Y] + \alpha^2\,\mathrm{Var}[Y]\right) \cdot \mathrm{Var}[Y]$$

$$= \mathrm{Var}[X] \cdot \mathrm{Var}[Y] - \left(\mathrm{Cov}[X,Y]\right)^2$$

□

Bemerkung 3.18. Es gilt Gleichheit in der Cauchy-Schwarz-Ungleichung g.d.w.

$$\text{es gibt } a, b, c \in \mathbb{R} \text{ (mit } a \neq 0 \text{ oder } b \neq 0\text{), so dass } P\big(aX + bY + c = 0\big) = 1.$$

In diesem Fall heißen X und Y *perfekt korreliert*.

Beispiel 3.19. 1. $X \sim \mathrm{Bin}_{n,p}$, schreibe $X = Y_1 + \dots + Y_n$ mit Y_i u.i.v. $\sim \mathrm{Ber}_p$, so ist (mit Satz 3.17, 2.)

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[Y_i] = n\mathrm{Var}[Y_1] = np(1-p)$$

(vgl. auch Bsp. 3.16, 3.).

2. $X \sim \mathrm{Hyp}_{s,w,n}$, so ist $\mathrm{Var}[X] = np(1-p)\left(1 - \frac{n-1}{s+w-1}\right)$

(s schwarze, w weiße Kugeln, ziehe n mal ohne Zurücklegen)

$A_i = \{i\text{-te gez. Kugel schwarz}\}$

$Y_i = \mathbb{1}_{A_i}$, $X = Y_1 + Y_2 + \dots + Y_n$,

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \underbrace{\mathrm{Var}[Y_i]}_{\mathrm{Var}[Y_1]} + \sum_{i \neq j}^{n} \underbrace{\mathrm{Cov}[Y_i, Y_j]}_{\mathrm{Cov}[Y_1, Y_2]}$$

$$\mathrm{Var}[Y_1] = \underbrace{\frac{s}{s+w}}_{=:p} \cdot \frac{w}{s+w}$$

$$\mathrm{Cov}[Y_1, Y_2] = \underbrace{\mathbb{E}[Y_1 \cdot Y_2]}_{\frac{s}{s+w} \cdot \frac{s-1}{s+w-1}} - \underbrace{\mathbb{E}[Y_1] \cdot \mathbb{E}[Y_2]}_{=\left(\frac{s}{s+w}\right)^2}$$

(vgl. Diskussion rund um Pólya-Urne)

$$= np(1-p) + n(n-1) \underbrace{\frac{s}{s+w}}_{=p} \underbrace{\left(\frac{s-1}{s+w-1} - \frac{s}{s+w} \right)}_{= -\frac{w}{s+w} \frac{1}{s+w-1}} = np(1-p)\left(1 - \frac{n-1}{s+w-1}\right)$$

3. Z reelle ZV mit $\mathbb{E}[\cancel{|X|^3}] < \infty$ und *symmetrischer* Verteilung, d.h. es gilt $P(Z > z) = P(Z < -z)$ für alle $z \geq 0$ (z.B. $Z \sim \mathcal{N}_{0,1}$), setze

$$Y := Z^2,$$

$\mathbb{E}[|Z|^3] < \infty$

$$\mathrm{Cov}[Y, Z] = \underbrace{\mathbb{E}[Z^2 \cdot Z]}_{= \mathbb{E}[Z^3] = 0} - \mathbb{E}[Z^2] \cdot \underbrace{\mathbb{E}[Z]}_{= 0} = 0$$

d.h. Y und Z sind unkorreliert

(aber i.A. <u>nicht</u> unabh.)

Definition 3.20. Seien $X, Y \in \mathcal{L}^2$.

wg. Cauchy-Schwarz-Ungl.

$$\kappa_{X,Y} := \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \in [-1,1]$$

heißt *Korrelationskoeffizient* von X und Y (manche Autoren schreiben auch $\rho_{X,Y}$).

Beobachtung 3.21 (Interpretation des Korrelationskoeffizienten via „beste lineare Vorhersage"). Es ist

$$\min_{\beta_0,\beta_1\in\mathbb{R}} \mathbb{E}\big[\underbrace{(Y-\beta_1 X-\beta_0)^2}_{(Y-(\beta_0+\beta_1 X))^2}\big] = (1-\kappa_{X,Y}^2)\min_{\beta_0\in\mathbb{R}}\mathbb{E}\big[(Y-\beta_0)^2\big] \quad \Big(= (1-\kappa_{X,Y}^2)\mathrm{Var}[Y]\Big),$$

(siehe Skript)

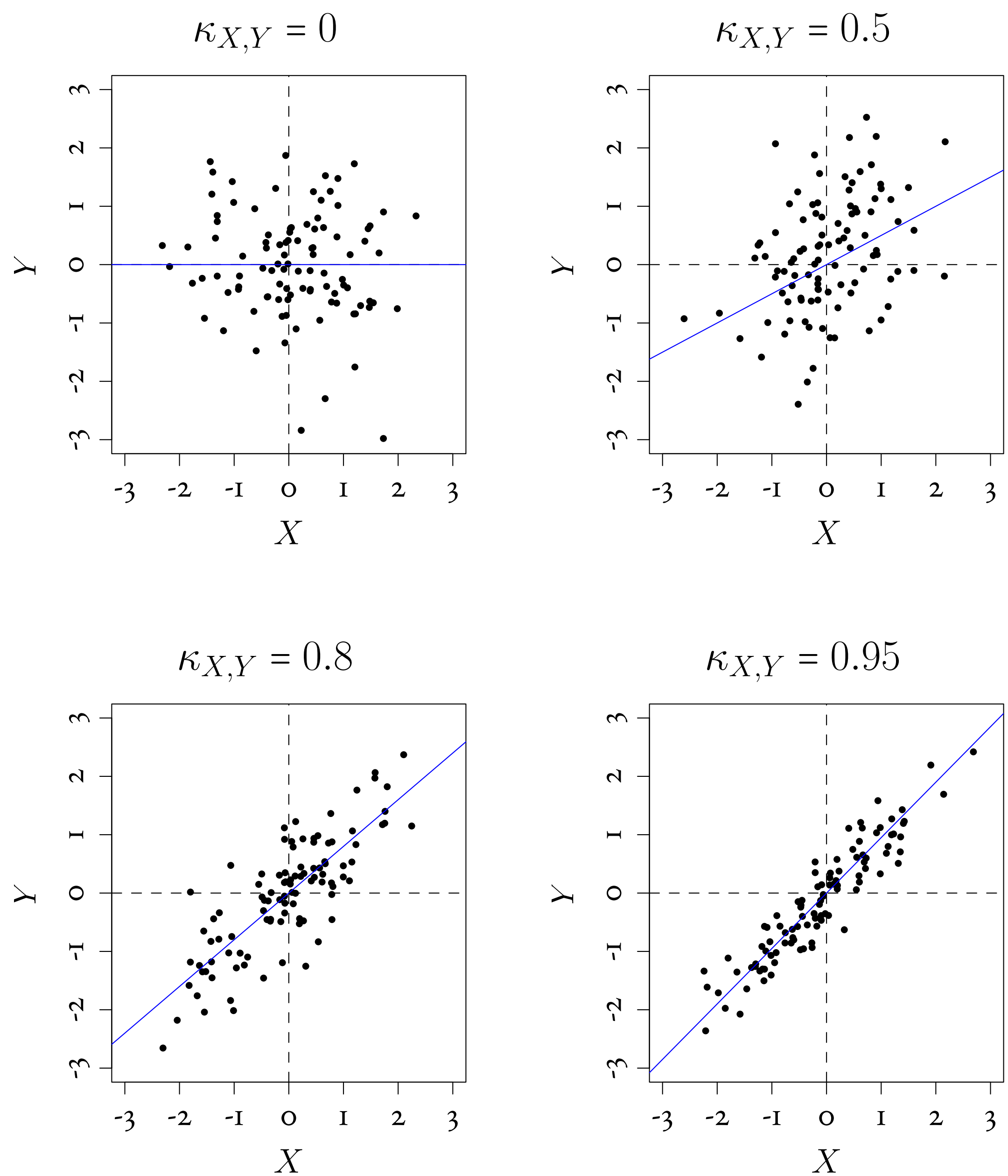
$$\min_{\beta_0,\beta_1\in\mathbb{R}} \mathbb{E}\big[(Y-\beta_1 X-\beta_0)^2\big] = (1-\kappa_{X,Y}^2)\min_{\beta_0\in\mathbb{R}} \mathbb{E}\big[(Y-\beta_0)^2\big] \quad \Big(= (1-\kappa_{X,Y}^2)\mathrm{Var}[Y]\Big),$$

Im Sinne einer möglichst kleinen quadratischen Abweichung ist $\mathbb{E}[Y]$ die beste konstante „Vorhersage“ von Y. Man kann demnach um einen Faktor $(1-\kappa_{X,Y}^2)$ besser vohersagen, wenn man stattdessen eine affin-lineare Funktion von X verwenden darf.

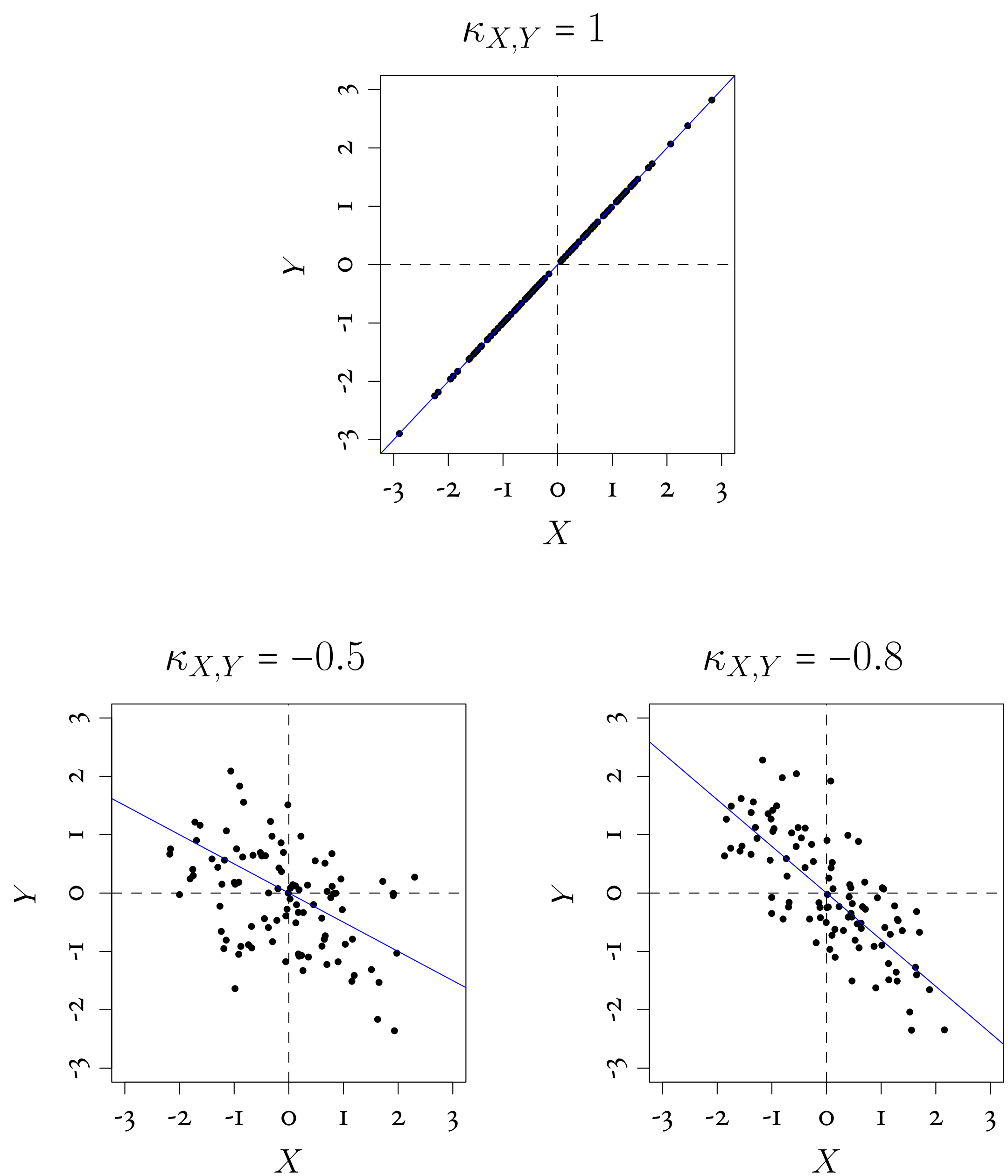
Demnach (vgl. auch Bem. 3.18)

$\lvert\kappa_{X,Y}\rvert = 1$	$\leftrightarrow$	perfekter linearer Zusammenhang zwischen X und Y
$\kappa_{X,Y} = 1$	$\leftrightarrow$	perfekter linearer Zusammenhang zwischen X und Y mit positivem Koeffizienten (X größer als $\mathbb{E}[X] \iff Y$ größer als $\mathbb{E}[Y]$)
$\kappa_{X,Y} = -1$	$\leftrightarrow$	perfekter linearer Zusammenhang zwischen X und Y mit negativem Koeffizienten (X größer als $\mathbb{E}[X] \iff Y$ kleiner als $\mathbb{E}[Y]$)

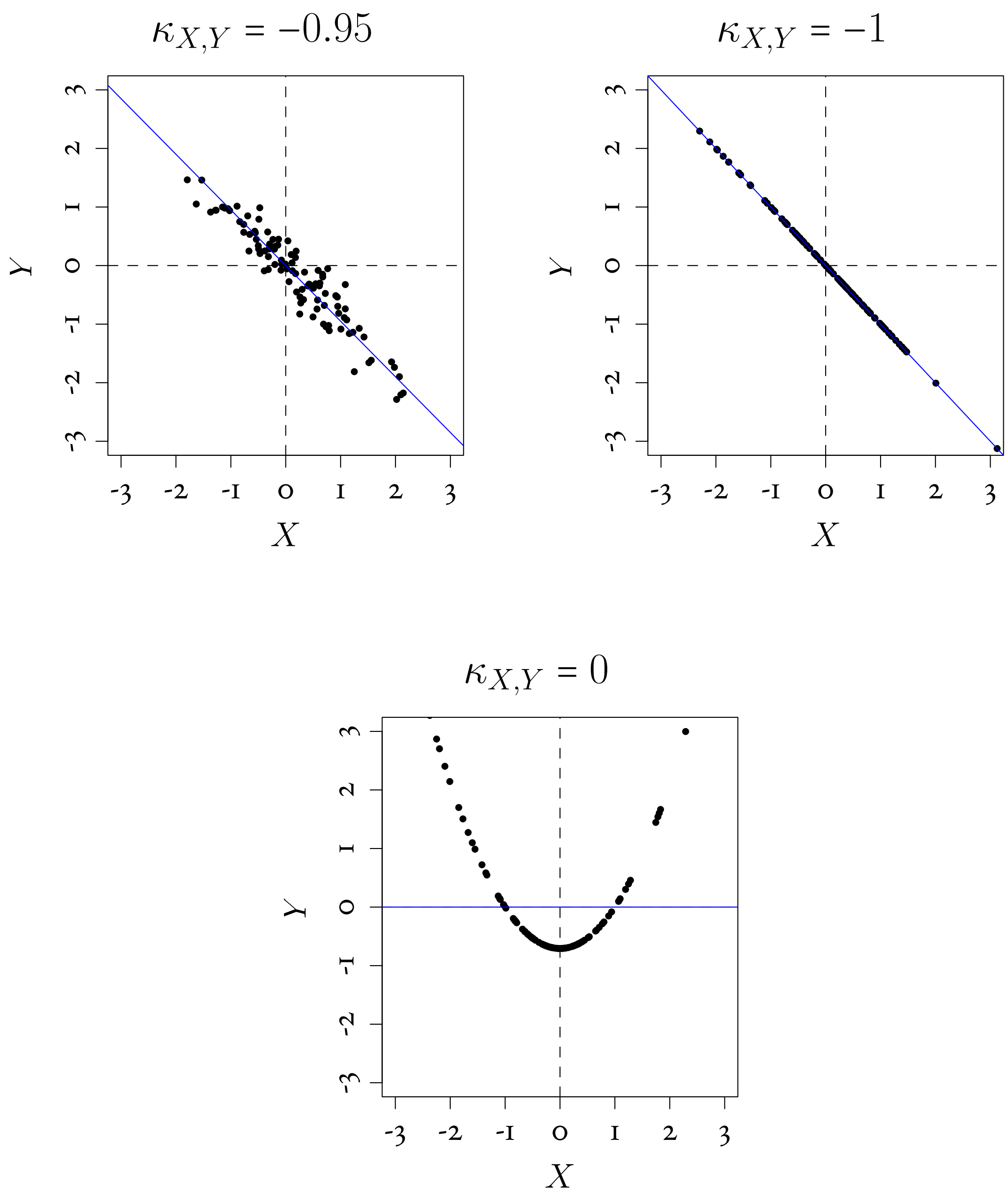
Nicht-lineare Zusammenhänge erfasst der Korrelationskoeffizient möglicherweise nicht korrekt (oder gar nicht), vgl. Bsp. 3.19, 3.



Die Scatterplots zeigen jeweils 100 simulierte Paare (X, Y), wobei $\sigma_X = \sigma_Y = 1$ und $\kappa_{X,Y}$ den angegebenen Wert hat. (Blau eingezeichnet ist die „Vorhersagegerade" $x \mapsto \beta_1^* x + \beta_0^*$.)



Die Scatterplots zeigen jeweils 100 simulierte Paare (X, Y), wobei $\sigma_X = \sigma_Y = 1$ und $\kappa_{X,Y}$ den angegebenen Wert hat. (Blau eingezeichnet ist die „Vorhersagegerade" $x \mapsto \beta_1^* x + \beta_0^*$.)



Die Scatterplots zeigen jeweils 100 simulierte Paare (X, Y), wobei $\sigma_X = \sigma_Y = 1$ und $\kappa_{X,Y}$ den angegebenen Wert hat. (Blau eingezeichnet ist die „Vorhersagegerade“ $x \mapsto \beta_1^* x + \beta_0^*$.)