WT Tutorium 6

Andreas Tsouchlos

30. Januar 2026

Erinnerung: Kovarianz, Korrelation & Korrelationskoeffizient

- Kovarianz

$$\mathrm{cov}(X,Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big)$$
$$= E(XY) - E(X)E(Y)$$

Erinnerung: Varianz

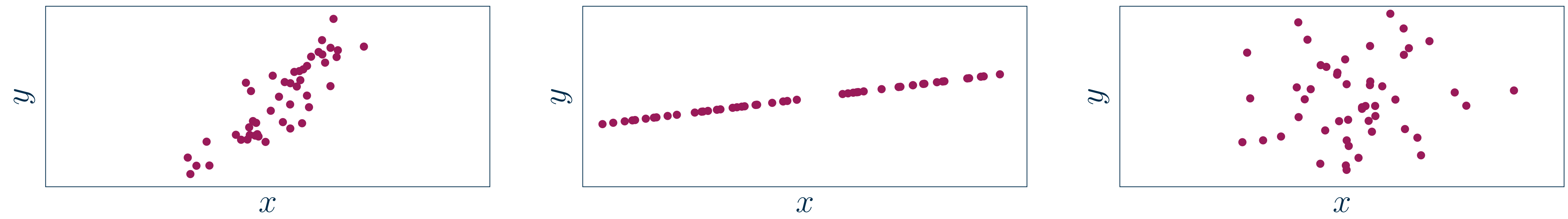
$$V(X) = E\Big(\big(X - E(X)\big)^2\Big) = E(X^2) - E^2(X)$$

- Korrelation

$$E(XY)$$

- Korrelationskoeffizient

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{V(X)V(Y)}}, \qquad \rho_{XY} \in [-1, 1] \qquad \rho_{XY} = 0 \;\Leftrightarrow\; E(XY) = E(X)E(Y)$$



$\rho = 0.9$ $\qquad$ $\rho = 1$ $\qquad$ $\rho = 0$

Unabhängigkeit, Korrelation & Disjunktheit

Disjunkte Ereignisse

"A und B können nicht gleichzeitig eintreten"

$$P(A \cap B) = 0$$
$$\Leftrightarrow$$
$$P(A \cup B) = P(A) + P(B)$$

Unabhängige Ereignisse

"Es gibt keine Zusammenhang zwischen dem Eintreten von A und B"

$$P(A \cap B) = P(A)P(B)$$
$$\overset{P(B)\neq 0}{\Leftrightarrow}$$
$$P(A|B) = P(A)$$

Unkorrelierte ZV

"Es gibt keinen linearen (genauer: affinen) Zusammenhang zwischen X und Y"

$$\rho_{XY} = 0$$
$$\Leftrightarrow$$
$$E(XY) = E(X)E(Y)$$

Unabhängige ZV

"Es gibt keinen Zusammenhang zwischen den Werten, welche X und Y annehmen"

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad \text{(stetig)}$$

bzw.

$$P_{X,Y}(x,y) = P_X(x)P_Y(y) \qquad \text{(diskret)}$$

Zusammenfassung

Korrelationskoeffizient

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

Kovarianz

$$\text{cov}(X,Y) = E(X,Y) - E(X)E(Y)$$

Erwartungswert

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

$$E(X+b) = E(X) + b$$
$$E(X+Y) = E(X) + E(Y)$$
$$E(aX) = aE(X)$$

Varianz

$$V(X) = E\big((X - E(X))^2\big)$$

$$V(X) = E(X^2) - (E(X))^2$$
$$V(aX) = a^2 V(X)$$
$$V(X+b) = V(X)$$

Aufgabe 1: Korrelationskoeffizienten

Es ist die Zufallsvariable $X \sim \mathcal{N}(0,1)$ gegeben. Berechnen Sie jeweils den Korrelationskoeffizienten ρ_{XY} für

a) $Y = aX + b \quad$ mit $a, b \in R$ und $a \neq 0$.

b) $Y = X^2$.

Aufgabe 1: Korrelationskoeffizienten

Es ist die Zufallsvariable $X \sim \mathcal{N}(0,1)$ gegeben. Berechnen Sie jeweils den Korrelationskoeffizienten ρ_{XY} für

a) $Y = aX + b$ mit $a, b \in R$ und $a \neq 0$.

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$\text{cov}(X,Y) = E(XY) - \cancelto{0}{E(X)}E(Y) = E(XY)$$

$$= E(aX^2 + bX) = a\underbrace{E(X^2)}_{=V(X)=1} + b\cancelto{0}{E(X)} = a$$

$$V(Y) = E\Big((Y - E(Y))^2\Big) = E\Big((aX)^2\Big) = a^2\underbrace{E(X^2)}_{=V(X)=1} = a^2$$

$$\rho_{XY} = \frac{a}{\sqrt{a^2}} = \frac{a}{|a|} = \begin{cases} +1, & a > 0 \\ -1, & a < 0 \end{cases}$$

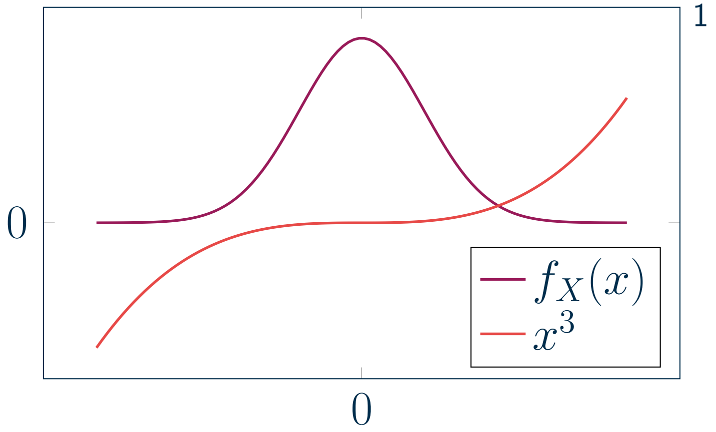
Aufgabe 1: Korrelationskoeffizienten

b) $Y = X^2$.

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

$$\text{cov}(X,Y) = E(XY) - \cancelto{0}{E(X)}E(Y) = E(XY) = E(X^3)$$
$$= \int_{-\infty}^{\infty} \underbrace{x^3}_{\text{ungerade}} \cdot \underbrace{f_X(x)}_{\text{gerade}} \, dx = 0$$

$$\rho_{XY} = 0$$



[1] Die zwei Kurven sind bezüglich der y-Achse unterschiedlich skaliert.

Erinnerung: Rechnen mit Normalverteilungen

- Die Normalverteilung

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(u-\mu)^2}{2\sigma^2}\right) du$$

- Die Standardnormalverteilung

$$X \sim \mathcal{N}(0,1)$$

$$\Phi(x) := F_X(x) = P(X \leq x)$$
$$\Phi(-x) = 1 - \Phi(x)$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,500000	0,10	0,539828	0,20	0,579260
0,02	0,507978	0,12	0,547758	0,22	0,587064
0,04	0,515953	0,14	0,555670	0,24	0,594835
0,06	0,523922	0,16	0,563559	0,26	0,602568
0,08	0,531881	0,18	0,571424	0,28	0,610261

- Standardisierung einer ZV

$$\widetilde{X} = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}$$

Rechenbeispiel

$$X \sim \mathcal{N}(\mu = 1, \sigma^2 = 0{,}5^2)$$

$$P\left(X \leq 1{,}12\right) = P\left(\frac{X-1}{0{,}5} \leq \frac{1{,}12-1}{0{,}5}\right) = P\Big(\underbrace{\widetilde{X}}_{\sim\mathcal{N}(0,1)} \leq 0{,}24\Big) = \Phi\left(0{,}24\right) = 0{,}594835$$

Grenzwertsätze

- Chintschin'sches Gesetz großer Zahlen

$$\left.\begin{array}{r} X_1, \dots X_N \text{ unabhängig und identisch verteilt} \\ E(X_1) < \infty \end{array}\right\} \Rightarrow \lim_{N \to \infty} P\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_N - E(X_1)\right| < \epsilon\right) = 1$$

"Je mehr realisierungen betrachtet werden, desto wahrscheinlicher ist das arithmetische Mittel nah am Erwartungswert"

- Zentraler Grenzwertsatz von Lindeberg-Lévy

$$\left.\begin{array}{r} X_1, \dots X_N \text{ unabhängig und identisch verteilt} \\ E(X_1) < \infty \\ V(X_1) < \infty \end{array}\right\} \Rightarrow \left\{\begin{array}{l} S_N = X_1 + \dots + X_N, \qquad a < b \in \mathbb{R} \\ \lim\limits_{N \to \infty} P\left(a \leq \frac{S_N - N\mu}{\sqrt{N\sigma^2}} \leq b\right) = \Phi(b) - \Phi(a) \end{array}\right.$$

"Die Summe unabhängiger und identisch verteilter ZV verhält sich immer mehr wie eine Normalverteilung, je mehr ZV betrachtet werden"

Approximation einer Binomialverteilung mit dem ZGWS

- Grenzwertsatz von de Moivre-Laplace

$$\lim_{N\to\infty} P\left(\frac{S_N - Np}{\sqrt{Np(1-p)}} \leq x\right) = \Phi(x)$$

$$\text{“Bin}(N,p) \xrightarrow{N\to\infty} \mathcal{N}\big(\mu = Np, \sigma^2 = Np(1-p)\big)\text{”}$$

Errinerung: Binomialverteilung

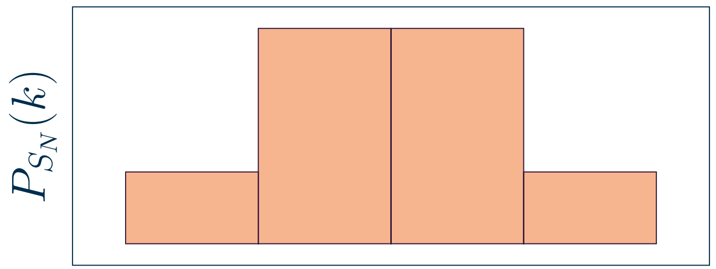
$$S_N \sim \text{Bin}(N,p)$$

$$P_{S_N}(k) = \binom{N}{k} p^k (1-p)^{N-k}$$

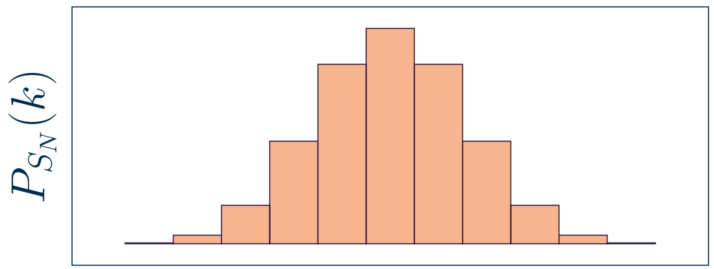
$$E(S_N) = Np, \quad V(S_N) = Np(1-p)$$

- Die approximation einer Binomialverteilung durch eine Normalverteilung ist in der Praxis dann zulässig, wenn $Np(1-p) \geq 9$:

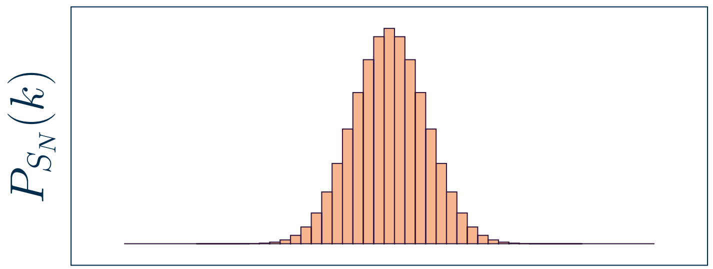
$$P_X(a < S_N \leq b) = \sum_{k=a}^{b} \binom{N}{k} p^k (1-p)^{N-k} \quad \approx \quad \Phi\left(\frac{b - Np}{\sqrt{Np(1-p)}}\right) - \Phi\left(\frac{a - Np}{\sqrt{Np(1-p)}}\right)$$



k
$N = 4, p = 0{,}5$



k
$N = 10, p = 0{,}5$



k
$N = 50, p = 0{,}5$

Zusammenfassung

Standardnormalverteilung

$$X \sim \mathcal{N}(0,1)$$

$$\Phi(x) := F_X(x) = P(X \leq x)$$
$$\Phi(-x) = 1 - \Phi(x)$$

Standardisierung

$$\widetilde{X} = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}$$

Approximation einer Binom.vert. mit dem ZGWS

Bedingung: $Np(1-p) \geq 9$

$$P_X(a < S_N \leq b) = \sum_{k=a}^{b} \binom{N}{k} p^k (1-p)^{N-k}$$
$$\approx \Phi\left(\frac{b-Np}{\sqrt{Np(1-p)}}\right) - \Phi\left(\frac{a-Np}{\sqrt{Np(1-p)}}\right)$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,60	0,945201	2,00	0,977250	2,40	0,991802
1,62	0,947384	2,02	0,978308	2,42	0,992240
1,64	0,949497	2,04	0,979325	2,44	0,992656
1,66	0,951543	2,06	0,980301	2,46	0,993053
1,68	0,953521	2,08	0,981237	2,48	0,993431

Aufgabe 2: Abschätzungen von Verteilungen (ZGWS)

Im Werk einer Zahnradfabrik werden verschiedene Präzisionsmetallteile gefertigt. Während einer Schicht werden 5000 Stück eines Typs A hergestellt. Bei der Qualitätskontrolle werden 3% dieser Teile als defekt klassifiziert und aussortiert.

a) Berechnen Sie näherungsweise die Wahrscheinlichkeit dafür, dass während einer Schicht zwischen 125 und 180 Teile aussortiert werden.

b) Die aussortierten Teile werden nach Schichtende zur Wiederverwertung in einem Kessel auf einmal eingeschmolzen. Wie viele Teile muss der Kessel fassen, damit er mit einer Wahrscheinlichkeit von min. $0{,}98$ nicht überfüllt ist?

c) Der Kessel fasse maximal 200 Teile. Es sollen nun mehr als 5000 Teile pro Schicht hergestellt werden. Wie viele Teile können maximal gefertigt werden, damit der Kessel mit einer Wahrscheinlichkeit von $0,98$ nicht überfüllt ist?

Aufgabe 2: Abschätzungen von Verteilungen (ZGWS)

Im Werk einer Zahnradfabrik werden verschiedene Präzisionsmetallteile gefertigt. Während einer Schicht werden 5000 Stück eines Typs A hergestellt. Bei der Qualitätskontrolle werden 3% dieser Teile als defekt klassifiziert und aussortiert.

a) Berechnen Sie näherungsweise die Wahrscheinlichkeit dafür, dass während einer Schicht zwischen 125 und 180 Teile aussortiert werden.

$$S_N := \text{Anzahl an defekten Teilen}$$

$$S_N \sim \text{Bin}(N = 5000, p = 0{,}03)$$

Viel zu aufwendig

$$P(125 \leq S_N \leq 180) = \sum_{k=125}^{180} \binom{N}{k} p^k (1-p)^{N-k}$$

$$Np(1-p) = 145{,}5 \geq 9 \qquad \rightarrow \qquad \tilde{S}_N \sim \mathcal{N}(\mu = \overbrace{Np}^{=150}, \sigma^2 = \overbrace{Np(1-p)}^{=145{,}5})$$

$$P(125 \leq \tilde{S}_N \leq 180) = P\left(\frac{125 - 150}{\sqrt{145{,}5}} \leq \overbrace{\frac{\tilde{S}_N - E(\tilde{S}_N)}{\sqrt{V(\tilde{S}_N)}}}^{\sim \mathcal{N}(0{,}1)} \leq \frac{180 - 150}{\sqrt{145{,}5}} \right)$$

$$\approx \Phi(2{,}487) - \Phi(-2{,}073)$$

$$= \Phi(2{,}487) - (1 - \Phi(2{,}073)) \approx 0{,}974$$

Aufgabe 2: Abschätzungen von Verteilungen (ZGWS)

b) Die aussortierten Teile werden nach Schichtende zur Wiederverwertung in einem Kessel auf einmal eingeschmolzen. Wie viele Teile muss der Kessel fassen, damit er mit einer Wahrscheinlichkeit von min. $0{,}98$ nicht überfüllt ist?

$$\tilde{S}_N \sim \mathcal{N}(\mu = 150, \sigma^2 = 145{,}5)$$

$$\begin{aligned}
P(\tilde{S}_N \leq dx) \geq 0{,}98 \quad &\Rightarrow \quad P\left(\overbrace{\frac{\tilde{S}_N - E(\tilde{S}_N)}{\sqrt{V(\tilde{S}_N)}}}^{\sim\mathcal{N}(0,1)} \leq \frac{dx - 150}{\sqrt{145{,}5}} \right) \geq 0{,}98 \\
&\Rightarrow \quad \Phi\left(\frac{dx - 150}{\sqrt{145{,}5}}\right) \geq 0{,}98 \\
&\Rightarrow \quad \frac{dx - 150}{\sqrt{145{,}5}} \geq \Phi^{-1}(0{,}98) = 2{,}06 \\
&\Rightarrow \quad dx \geq 174{,}8
\end{aligned}$$

Der Kessel muss mindestens 175 Teile fassen

Aufgabe 2: Abschätzungen von Verteilungen (ZGWS)

c) Der Kessel fasse maximal 200 Teile. Es sollen nun mehr als 5000 Teile pro Schicht hergestellt werden. Wie viele Teile können maximal gefertigt werden, damit der Kessel mit einer Wahrscheinlichkeit von $0{,}98$ nicht überfüllt ist?

$$\tilde{S}_N \sim \mathcal{N}\left(\mu = Np, \sigma^2 = Np(1-p)\right)$$

$$\begin{aligned}
P(\tilde{S}_N \leq 200) \geq 0{,}98 \quad &\Rightarrow \quad P\left(\overbrace{\frac{\tilde{S}_N - E(\tilde{S}_N)}{\sqrt{V(\tilde{S}_N)}}}^{\sim\mathcal{N}(0,1)} \leq \frac{200 - Np}{\sqrt{Np(1-p)}}\right) \geq 0{,}98 \\
&\Rightarrow \quad \Phi\left(\frac{200 - Np}{\sqrt{Np(1-p)}}\right) \geq 0{,}98 \\
&\Rightarrow \quad \frac{200 - Np}{\sqrt{Np(1-p)}} \geq \Phi^{-1}(0{,}98) = 2{,}06 \\
&\Rightarrow \quad Np + 2{,}06 \cdot \sqrt{Np(1-p)} - 200 \leq 0
\end{aligned}$$

$$\begin{aligned}
u &:= \sqrt{N} \\
a &:= p = 0{,}03 \\
b &:= 2{,}06 \cdot \sqrt{p(1-p)} \approx 0{,}351 \\
c &:= -200
\end{aligned}$$

$$\rightarrow \quad au^2 + bu + c \leq 0$$

$$\begin{aligned}
&\Rightarrow \quad u = \sqrt{N} \in [-76, 87{,}7] \\
&\Rightarrow \quad N \leq 5776 \quad \cap \quad N \leq 7691{,}29 \\
&\Rightarrow \quad N \leq 5776
\end{aligned}$$