Data Science 2

Prof. Dr. Mark Trede

Institut für Ökonometrie und Wirtschaftsstatistik

September 2023

Zufallsvektoren:

Für viele interessante Anwendungen braucht man mehrere Zufallsvariablen (Vereinfachung: $n = 2$)

Beispiele:

- X: Haushaltsgröße, Y: Anzahl Autos
- X: Aktienrendite Siemens, Y: Aktienrendite Zalando
- X: kleinere Augenzahl, Y: größere Augenzahl von zwei Würfeln

Gemeinsame Verteilungsfunktion

Seien $X : \Omega \longrightarrow \mathbb{R}$ und $Y : \Omega \longrightarrow \mathbb{R}$ zwei Zufallsvariablen. Dann ist

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

die gemeinsame Verteilungsfunktion von X und Y

Eigenschaften der gemeinsamen Verteilungsfunktion:

- $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ ist monoton steigend in x und y
- Es gilt $F_{X,Y}(-\infty, y) = F_{XY}(x, -\infty) = 0$
- Es gilt $F_{X,Y}(+\infty, +\infty) = 1$

Gemeinsam diskrete Zufallsvariablen

X und Y heißen gemeinsam diskret, wenn es endlich viele oder abzählbar unendlich viele Werte $x_{1,} x_{2,}$ ... und $y_{1,} y_{2,}$ ... gibt, so dass

$$\sum_j \sum_k p_{jk} = 1$$

mit $p_{jk} = P(X = x_j, Y = y_k)$

Zufallsvektoren

Gemeinsam diskrete Zufallsvariablen

- Gemeinsame Wahrscheinlichkeitsfunktion

$$f_{X,Y}(x,y) = \begin{cases} p_{jk} & \text{für } x = x_j \text{ und } y = y_k \\ 0 & \text{sonst} \end{cases}$$

- Darstellung als Wahrscheinlichkeitstabelle

$X \backslash Y$	y_1	...	y_K
x_1	p_{11}	...	p_{1K}
$\vdots$	$\vdots$		$\vdots$
x_J	p_{J1}	...	p_{JK}

Beispiel: Haushalte

Sei X: Haushaltsgröße; Y: Anzahl Autos

$X \backslash Y$	0	1	2
1	0.10	0.14	0.01
2	0.05	0.15	0.10
3	0.02	0.10	0.08
4	0.02	0.06	0.07
5	0.01	0.05	0.04

Wie groß ist $F_{X,Y}(3,1)$?

Beispiel: Haushalte

Sei X: Haushaltsgröße; Y: Anzahl Autos

$X\backslash Y$	0	1	2
1	0.10	0.14	0.01
2	0.05	0.15	0.10
3	0.02	0.10	0.08
4	0.02	0.06	0.07
5	0.01	0.05	0.04

Wie groß ist $F_{X,Y}(3,1)$? $P(X \leq 3, Y \leq 1) = 0.56$

Beispiel: Haushalte

Sei X: Haushaltsgröße; Y: Anzahl Autos

$X\backslash Y$	0	1	2
1	0.10	0.14	0.01
2	0.05	0.15	0.10
3	0.02	0.10	0.08
4	0.02	0.06	0.07
5	0.01	0.05	0.04

Wie groß ist $F_{XY}(1.5, 3.2)$?

Beispiel: Haushalte

Sei X: Haushaltsgröße; Y: Anzahl Autos

$X\backslash Y$	0	1	2
1	0.10	0.14	0.01
2	0.05	0.15	0.10
3	0.02	0.10	0.08
4	0.02	0.06	0.07
5	0.01	0.05	0.04

Wie groß ist $F_{XY}(1.5, 3.2)$? $P(X \leq 1.5, Y \leq 3.2) = 0.25$

Gemeinsam stetige Zufallsvariablen

X und Y heißen gemeinsam stetig, falls es eine Funktion $f_{X,Y}$ gibt mit

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v) du dv$$

Die Funktion $f_{X,Y}$ heißt gemeinsame Dichte (Dichtefunktion)

Zufallsvektoren

- Bei partieller Differenzierbarkeit gilt

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

- Für $x, y \in \mathbb{R}$ gilt $f_{X,Y}(x,y) \geq 0$
- Das Volumen unter der Dichte ist

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$$

Zufallsvektoren

- Durch Doppel-Integrale der Dichte erhält man Wahrscheinlichkeiten, so ist beispielsweise

$$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2)$$
$$= \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{XY}(x, y) dx dy$$

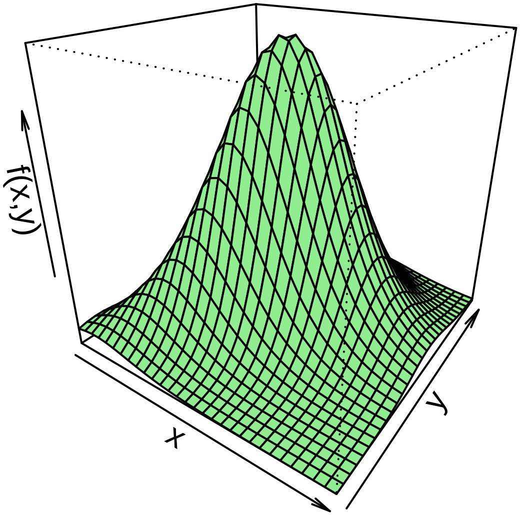
- Wegen der Stetigkeit spielt es keine Rolle, ob die Ungleichungen strikt sind oder nicht

Zufallsvektoren

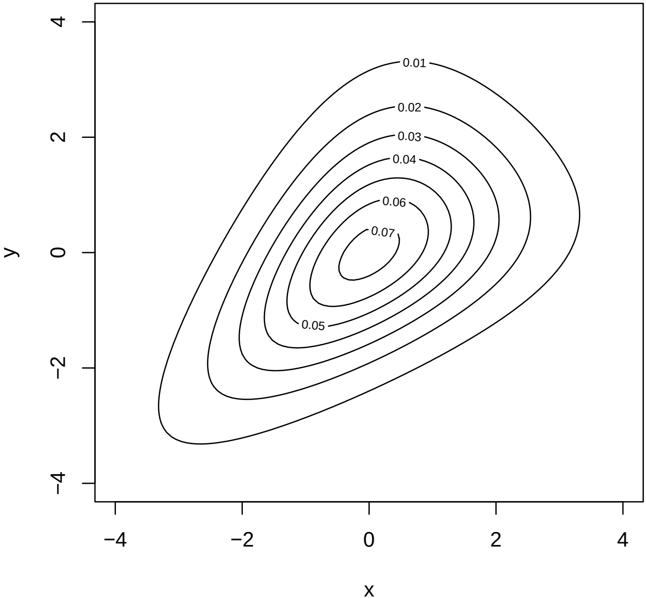
Visualisierung einer gemeinsamen Dichte:

- Plot als 3D-"Gebirge"
- Höhenlinien (Contour-Plot)
- Farben als dritte Dimension (Image-Plot)
- Beispiel:
$$f_{X,Y}(x,y) = \frac{2e^{-x-y}}{(1+e^{-x}+e^{-y})^3}$$

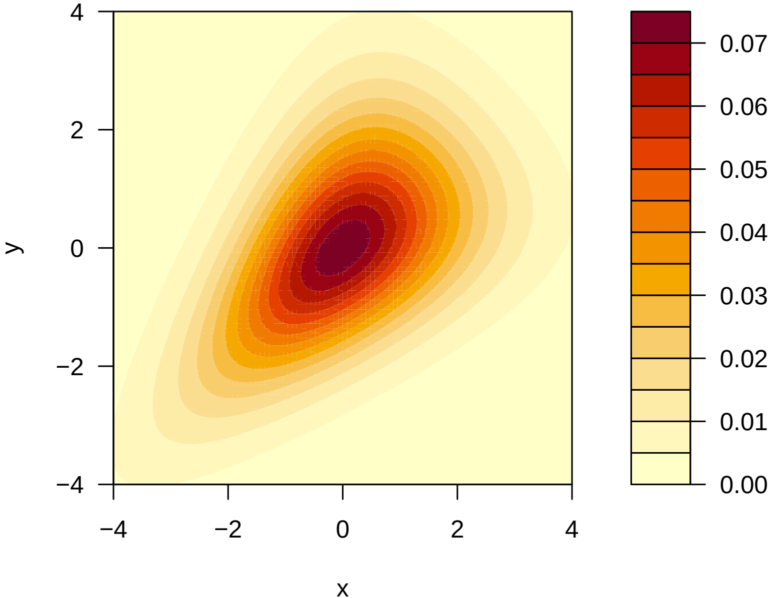
Zufallsvektoren



Zufallsvektoren



Zufallsvektoren



Randverteilungen

- Wie ist die eine Zufallsvariable verteilt, wenn man die andere ignoriert?
- Randverteilungsfunktionen F_X und F_Y

$$F_X(x) = F_{X,Y}(x, \infty) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

$$F_Y(y) = F_{X,Y}(\infty, y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

Randverteilungen gemeinsam diskreter Zufallsvariablen

$$\begin{aligned} p_{j\cdot} &= P(X = x_j) = \sum_k p_{jk} \\ p_{\cdot k} &= P(Y = y_k) = \sum_j p_{jk} \end{aligned}$$

Randdichten gemeinsam stetiger Zufallsvariablen

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

(die jeweils andere Variable wird “rausintegriert”)

Unabhängigkeit

Zwei Zufallsvariablen X und Y heißen (stochastisch) unabhängig, falls für alle $x, y \in \mathbb{R}$

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$

Randverteilungen F_X und F_Y ergeben sich aus der gemeinsamen Verteilung $F_{X,Y}$, aber $F_{X,Y}$ ist im Allgemeinen nicht eindeutig durch F_X und F_Y bestimmt

Zufallsvektoren
Unabhängigkeit

- Gemeinsam diskrete X und Y sind unabhängig, wenn für alle $j = 1, 2, \ldots$und $k = 1, 2, \ldots$

$$P(X = x_j, Y = y_k) = P(X = x_j) \cdot P(Y = y_k)$$

- Gemeinsam stetige X und Y sind unabhängig, wenn für alle $x, y \in \mathbb{R}$

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

Beispiel: Gemeinsam diskrete Verteilung

Zwei Würfel werden geworfen, X: kleinere Augenzahl, Y: größere Augenzahl

$X\backslash Y$	1	2	3	4	5	6	f_X
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{11}{36}$
2	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{9}{36}$
3	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{7}{36}$
4	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{5}{36}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$
f_Y	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1

Beispiel: Gemeinsam stetige Verteilung

Sei (X, Y) gemeinsam stetig verteilt

$$f_{X,Y}(x,y) = \begin{cases} x + y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Randdichte von X:

Beispiel: Gemeinsam stetige Verteilung

Sei (X, Y) gemeinsam stetig verteilt

$$f_{X,Y}(x,y) = \begin{cases} x + y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Randdichte von X:

$$\begin{aligned} f_X(x) &= \int_0^1 f_{X,Y}(x,y)dy = \int_0^1 (x+y)dy \\ &= x + \int_0^1 ydy = x + [0.5y^2]_0^1 = x + \frac{1}{2} \end{aligned}$$

Beispiel: Gemeinsam stetige Verteilung

Sei (X, Y) gemeinsam stetig verteilt

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

X und Y sind abhängig, weil:

Beispiel: Gemeinsam stetige Verteilung

Sei (X, Y) gemeinsam stetig verteilt

$$f_{X,Y}(x, y) = \begin{cases} x + y & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

X und Y sind abhängig, weil:

$$f_X(0.2) \cdot f_Y(0.2) = (0.2 + 0.5)(0.2 + 0.5) = 0.49$$
$$f_{XY}(0.2, 0.2) = 0.2 + 0.2 = 0.4$$