

Statistik für Informatiker, SS 2019

1 Grundlagen aus der Wahrscheinlichkeitstheorie

1.3 Bedingte Wahrscheinlichkeiten, Unabhängigkeit, gemeinsame Verteilung

1.3.4 Die Situation im Fall mit Dichten

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/StatInfo19/>



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

20.5.2018

Marginaldichten

Beobachtung 1.55 (Marginaldichten)

Die Zufallsvariable $X = (X_1, X_2)$ mit Werten in $(S \subset) \mathbb{R}^2$ habe (gemeinsame) Dichte $f_X(x_1, x_2)$, so hat X_1 die Dichte

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2 \quad (\text{für } x_1 \in \mathbb{R})$$

und analog hat X_2 die Dichte $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_1$ ($x_2 \in \mathbb{R}$).

f_{X_1} und f_{X_2} heißen die Marginal- oder Randdichten von f_X .

Argument:

$$P(X_1 \in [a, b]) = P(X_1 \in [a, b], X_2 \in \mathbb{R})$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_{[a,b]}(x_1) \mathbf{1}_{\mathbb{R}}(x_2) f_X(x_1, x_2) dx_1 dx_2 = \int_a^b \left(\int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2 \right) dx_1$$

Beobachtung 1.55 (Marginaldichten, allgemein)

Allgemein: Für $X = (X_1, \dots, X_d)$ mit Werten in \mathbb{R}^n und Dichte f_X ist

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

die i -te Marginaldichte.

Das kontinuierliche Analogon zu Prop. 1.52 lautet:

Bericht 1.56 (Unabhängigkeit im reellwertigen Fall mit Dichte)

Seien X_1, X_2, \dots, X_n reellwertige ZVn, $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}_+$ Wahrscheinlichkeitsdichten (d.h. $\int_{-\infty}^{\infty} f_i(x) dx = 1$), dann sind äquivalent:

- 1 X_1, \dots, X_n sind u.a. und X_i hat Dichte f_i für $i = 1, \dots, n$ (d.h. $P(X_i \in B) = \int_B f_i(x) dx$ und $P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$ für $B_1, \dots, B_n \subset \mathbb{R}$).
- 2 Die ZV $X = (X_1, \dots, X_n)$ mit Werten in \mathbb{R}^n hat Dichte

$$f(x) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

d.h. die gemeinsame Dichte hat Produktgestalt.

X_1, \dots, X_n unabhängig und X_i hat Dichte f_i



$X = (X_1, \dots, X_n)$ hat Dichte

$$f(x) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

Vergleicht man dies mit Beob. 1.55, so folgt:

In diesem Fall ist die gemeinsame Dichtefunktion das Produkt der Marginaldichten.

Beweisidee. „Naiv“ rechnen wir

$$\begin{aligned} & \int_{-\infty}^{x_1} f_1(y_1) dy_1 \cdots \int_{-\infty}^{x_n} f_n(y_n) dy_n \\ &= \int_{(-\infty, x_1] \times \cdots \times (-\infty, x_n]} f_1(x_1) \cdots f_n(x_n) dy_1 \dots dy_n \end{aligned}$$



Beispiel 1.57

1. Wähle $A = [a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$ in Bsp. 1.36, 1., d.h. $X = (X_1, X_2)$ ist uniform verteilt auf einem (achsenparallelen) Rechteck A (mit Fläche $\text{vol}(A) = (b_1 - a_1)(b_2 - a_2)$).

X hat Dichte

$$\begin{aligned} f_X(x_1, x_2) &= \frac{1}{(b_1 - a_1)(b_2 - a_2)} \mathbf{1}_{[a_1, b_1] \times [a_2, b_2]}(x_1, x_2) \\ &= f_{X_1}(x_1) f_{X_2}(x_2) \end{aligned}$$

mit Marginaldichten $f_{X_i}(x_i) = \frac{1}{(b_i - a_i)} \mathbf{1}_{[a_i, b_i]}(x_i) \quad (i = 1, 2)$,

d.h. die Koordinaten X_1 und X_2 sind unabhängig (und jeweils uniform auf $[a_i, b_i]$ verteilt).

Beispiel 1.57 (Forts.)

2. Wähle $A = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ in Bsp. 1.36, d.h. $X = (X_1, X_2)$ ist uniform verteilt auf dem Einheitskreis (mit Fläche $\text{vol}(A) = \pi$) mit Dichte

$$f_X(x_1, x_2) = \frac{1}{\pi} \mathbf{1}_{[0,1]}(x_1^2 + x_2^2).$$

Die Marginaldichte ist

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \frac{1}{\pi} \mathbf{1}_{[0,1]}(x_1^2 + x_2^2) dx_2 = \frac{1}{\pi} \mathbf{1}_{[-1,1]}(x_1) \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} 1 dx_2 \\ &= \mathbf{1}_{[-1,1]}(x_1) \frac{2}{\pi} \sqrt{1-x_1^2} \end{aligned}$$

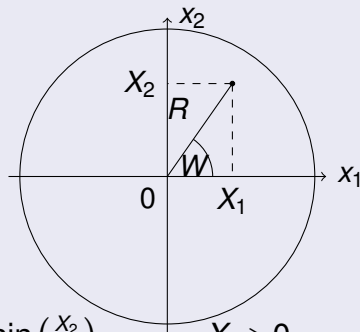
(und analog bzw. offensichtlich aus Symmetrie ist $f_{X_2} = f_{X_1}$).

Insbesondere sind (im Gegensatz zu 1.) X_1 und X_2 sind (natürlich) *nicht* unabhängig, denn $f_X(x_1, x_2) \neq f_{X_1}(x_1)f_{X_2}(x_2)$.

Beispiel 1.57 (Forts.)

3. Betrachten wir
 $X = (X_1, X_2)$ aus 2. in
 Polarkoordinaten:

Sei R der Radius, W
 der Winkel von X



also

$$R = \sqrt{X_1^2 + X_2^2}, \quad W = \begin{cases} \arcsin\left(\frac{X_2}{R}\right), & X_1 \geq 0, \\ \pi - \arcsin\left(\frac{X_2}{R}\right), & X_1 < 0, X_2 \geq 0, \\ -\pi - \arcsin\left(\frac{X_2}{R}\right), & X_1 < 0, X_2 < 0, \end{cases}$$

$$[\text{knapp} : W = \text{sign}(X_1) \cdot \arcsin\left(\frac{X_2}{R}\right) + \pi 1_{[-1,0)}(X_1) \cdot \text{sign}(X_2)]$$

bzw. $X_1 = R \cos(W)$, $X_2 = R \sin(W)$

Beispiel 1.57 (Forts.)

3. (X_1, X_2) uniform im Einheitskreis, in Polarkoordinaten:

$$R = \sqrt{X_1^2 + X_2^2}, \quad W = \text{sign}(X_1) \cdot \arcsin\left(\frac{X_2}{R}\right) + \pi \mathbf{1}_{[-1,0)}(X_1) \cdot \text{sign}(X_2)$$

Dann sind R und W
unabhängig,

R hat Dichte

$$f_R(r) = 2r \mathbf{1}_{[0,1]}(r),$$

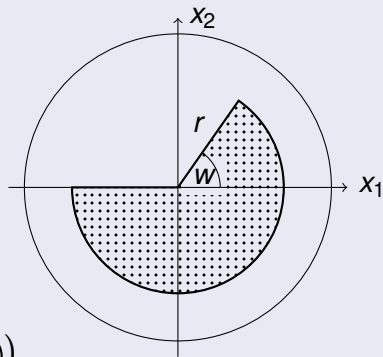
W hat Dichte

$$f_W(w) = \frac{1}{2\pi} \mathbf{1}_{[-\pi, \pi)}(w)$$

denn (für $0 \leq r \leq 1$, $-\pi \leq w < \pi$)

$$P(R \leq r, W \leq w) = P(X \in \text{shaded region})$$

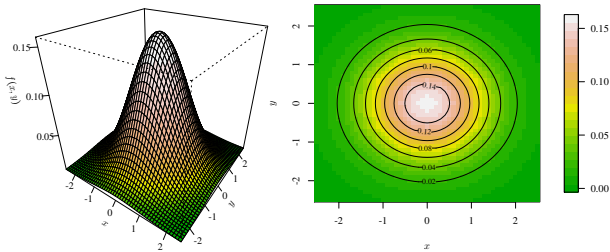
$$= \frac{\pi r^2 \frac{w+\pi}{2\pi}}{\pi 1^2} = r^2 \frac{w+\pi}{2\pi} = \int_0^r 2s \, ds \cdot \int_{-\pi}^w \frac{1}{2\pi} \, dv.$$



Beispiel 1.57 (Forts.)

4. $X = (X_1, \dots, X_d)$ d -dimensional
 Standard-normalverteilt, so sind X_1, \dots, X_d unabhängig
 und jeweils $\sim \mathcal{N}_{0,1}$
 (d.h. die X_i sind [1-dimensional] Standard-normalverteilt),
 denn

$$\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_d^2)\right) = \prod_{i=1}^d \left(\frac{1}{(2\pi)^{1/2}} e^{-x_i^2/2}\right)$$



Die Beobachtung zur Rotationssymmetrie aus
Bsp. 1.57, 3. verallgemeinert sich folgendermaßen:

Beobachtung 1.58

$X = (X_1, X_2)^T$ habe eine rotationssymmetrische Dichte f_X ,
d.h. $f_X(x_1, x_2)$ hängt nur von $r = \sqrt{x_1^2 + x_2^2}$ ab (also
 $f_X(x_1, x_2) = g(r)$ für eine gewisse Funktion $g \geq 0$),
 $X' = (X'_1, X'_2)^T$ entstehe aus X durch Drehung (um Winkel
 α um den Ursprung), d.h.

$$\begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \cos(\alpha)X_1 - \sin(\alpha)X_2 \\ \sin(\alpha)X_1 + \cos(\alpha)X_2 \end{pmatrix}.$$

Dann hat X' dieselbe Dichte (und somit dieselbe
Verteilung) wie X .

(Dies ist anschaulich sehr plausibel, man kann es z.B. mit Bericht 1.59 [unten] beweisen.)

Beobachtung 1.58 (Forts.)

$X = (X_1, X_2)^T$ habe eine rotationssymmetrische Dichte
 $f_X(x_1, x_2) = g(\sqrt{x_1^2 + x_2^2})$

Sei (R, W) die Polarkoordinatendarstellung von X (wie in Bsp. 1.57, 3., insbes. $R = \sqrt{X_1^2 + X_2^2}$), so sind R und W unabhängig, W ist uniform verteilt auf $[-\pi, \pi)$ und R hat Dichte $2\pi r g(r) 1_{[0, \infty)}(r)$:

$$\begin{aligned} P(R \leq u) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0, u]}(\sqrt{x_1^2 + x_2^2}) g(\sqrt{x_1^2 + x_2^2}) dx_1 dx_2 \\ &= \int_{-\pi}^{\pi} \int_0^u g(r) r dr dw = \int_0^u 2\pi r g(r) dr \end{aligned}$$

(wir verwenden hier etwas salopp die „Polarkoordinatenform des Flächenelements“ $dx_1 dx_2 = r dr dw$, man kann dies wiederum mit Bericht 1.59 [unten] beweisen.)

Speziell für X_1, X_2 unabhängig, $\sim \mathcal{N}_{0,1}$, somit $X = (X_1, X_2)$ 2-dim. Standard-normalverteilt, ergibt sich (für $u > 0$):

$$\begin{aligned} P(R^2 \leq u) &= P(R \leq \sqrt{u}) = \int_0^{\sqrt{u}} 2\pi r \frac{1}{2\pi} e^{-r^2/2} dr \\ &= \left[-e^{-r^2/2} \right]_{r=0}^{r=\sqrt{u}} = 1 - e^{-u/2} \end{aligned}$$

d.h. $R^2 \sim \text{Exp}_{1/2}$.

Dies ist der theoretische Hintergrund der *Box-Muller-Methode* zur Simulation normalverteilter ZVn.



Allgemeine Dichtetransformation im \mathbb{R}^d

Bericht 1.59 (Allgemeine Dichtetransformation im \mathbb{R}^d)

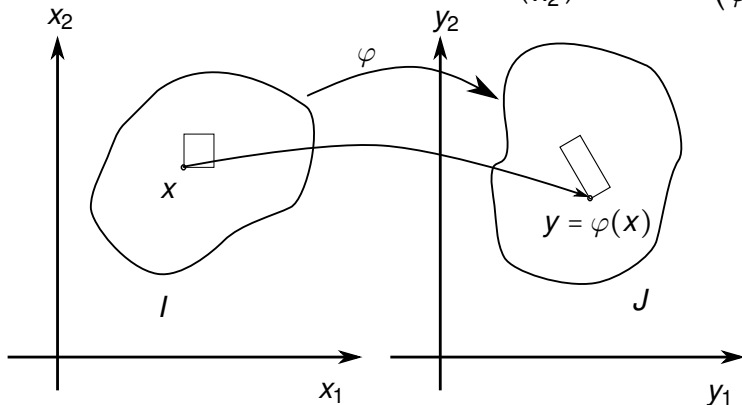
X \mathbb{R}^d -wertige ZV mit Dichte f_X , $I \subset \mathbb{R}^d$ offen mit $P(X \in I) = 1$, $J \subset \mathbb{R}^d$ offen, $\varphi : I \rightarrow J$ bijektiv, stetig differenzierbar mit Ableitung

$$\varphi'(x) = \left(\frac{\partial \varphi_i}{\partial x_j}(x) \right)_{i,j=1}^d \quad (\text{„Jacobi-Matrix“})$$

(wobei $\varphi(x) = (\varphi_1(x), \dots, \varphi_d(x))^T$, d.h. φ_i ist die i -te Koordinatenfunktion von φ), dann hat $Y := \varphi(X)$ die Dichte

$$f_Y(y) = \begin{cases} \frac{f_X(\varphi^{-1}(y))}{|\det \varphi'(\varphi^{-1}(y))|}, & y \in J, \\ 0, & y \notin J. \end{cases}$$

Heuristik (im Fall $d = 2$): Lokal sieht $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \end{pmatrix}$

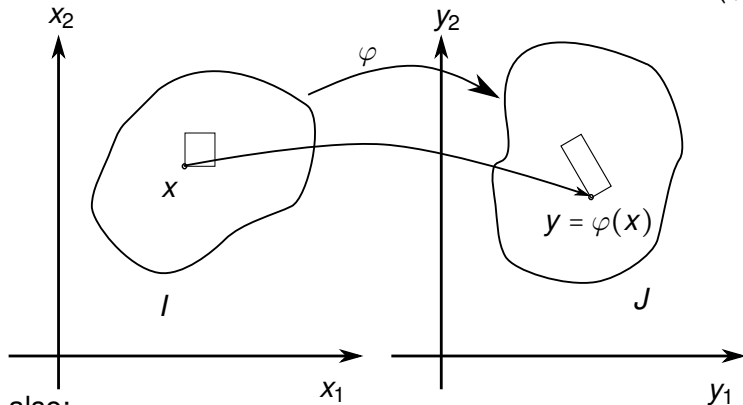


„aus wie“ $\varphi(x') \approx \varphi(x) + \varphi'(x) \cdot (x' - x)$

$$= \varphi(x) + \begin{pmatrix} \frac{\partial}{\partial x_1} \varphi_1(x) & \frac{\partial}{\partial x_2} \varphi_1(x) \\ \frac{\partial}{\partial x_1} \varphi_2(x) & \frac{\partial}{\partial x_2} \varphi_2(x) \end{pmatrix} \cdot \begin{pmatrix} x'_1 - x_1 \\ x'_2 - x_2 \end{pmatrix}$$

(plus Terme, die $O(\|x' - x\|^2)$ sind)

Heuristik (im Fall $d = 2$): Lokal sieht $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \end{pmatrix}$

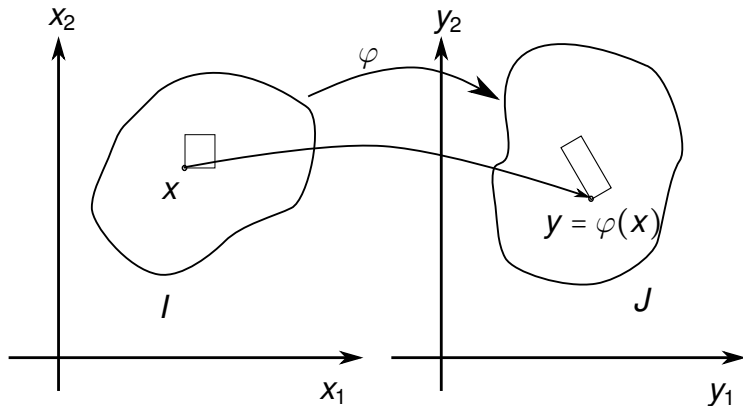


also:

die Fläche der Größe $h_1 \cdot h_2$ „rund um x “

wird auf

\approx Fläche $h_1 \cdot h_2 \cdot |\det \varphi'(x)|$ „rund um y “ abgebildet.



Wenden wir dies auf $Y = \varphi(X)$ an, so bedeutet das anschaulich:
 Für $y = \varphi(x) \in J$ (und sehr kleines $h > 0$) ist

$$\begin{aligned}
 f_Y(y)h^2 &\approx \mathbb{P}(Y \text{ nimmt Wert in Quadrat der Fläche } h^2 \text{ mit „Aufpunkt“ } y \text{ an}) \\
 &\approx \mathbb{P}(X \text{ nimmt Wert in Quader d. Fl. } h^2/|\det \varphi'(x)| \text{ mit „Aufpkt.“ } x \text{ an}) \\
 &\approx f_X(x) \frac{h^2}{|\det \varphi'(x)|} = \frac{f_X(\varphi^{-1}(y))}{|\det \varphi'(\varphi^{-1}(y))|} h^2.
 \end{aligned}$$



d -dimensionale Verteilungsfunktionen

Nur der Vollständigkeit halber, wir werden dies im Verlauf der Vorlesung nicht benötigen:

Bericht 1.60

Analog betrachtet zu Def. 1.26 betrachtet man im Fall $d > 1$ für $(x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ gelegentlich die d -dimensionale Verteilungsfunktion

$$F_X(x_1, x_2, \dots, x_d) := P(X \in (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]),$$

die allerdings etwas weniger „handlich“ ist als im 1-dimensionalen Fall.

Bericht 1.60 (Forts.)

Analog zu Bem. 1.27, 3. „weiß“ die d -dimensionale Verteilungsfunktion von X „alles“ über die Verteilung von X .

Eigenschaften:

- 1 F_X rechtsstetig, d.h. $x_n \searrow x$ (koordinatenweise) $\Rightarrow F_X(x_n) \rightarrow F(x)$
- 2 $F_X(x_n) \rightarrow 1$ wenn $x_n \rightarrow (+\infty, \dots, +\infty)$
- 3 $F_X(x_n) \rightarrow 0$ wenn $\min_{i=1, \dots, d} x_{n,i} \rightarrow -\infty$
- 4 Für $x = (x_1, \dots, x_d) < y = (y_1, \dots, y_d)$ (koordinatenweise), parametrisiere die Ecken des d -Quaders $(x, y]$ mit $\{1, 2\}^d$ via $\{u = (z_1^{(i_1)}, \dots, z_d^{(i_d)}) : i_1, \dots, i_d \in \{1, 2\}\}$ wo $z_j^{(1)} = x_j$, $z_j^{(2)} = y_j$, es muss gelten

$$\sum_{u \text{ Ecken}} (-1)^{\#\{1 \leq m \leq d : i_m = 1\}} F_X(u) \left(= \mu_F((x, y]) \right) \geq 0$$