

Kapitel 3

Verteilungen mehrerer Variablen

3.1 Eigenschaften von Verteilungen mehrerer Variablen

Im allgemeinen muss man Wahrscheinlichkeiten für mehrere Variable, die häufig auch voneinander abhängen, gleichzeitig betrachten.

Beispiele:

- Wir hatten im letzten Kapitel bereits die Multinomial-Verteilung als Beispiel einer Verteilung, die von mehreren diskreten Variablen abhängt, kennengelernt.
- Die Dichte einer Ladungswolke um eine Glühkathode hat eine dreidimensionale Verteilung.
- Ein System von n Teilchen hat eine Wahrscheinlichkeitsdichte in dem $6n$ -dimensionalen Orts-Impulsraum (= Phasenraum). Zum Beispiel sind für ein ideales Gas die Ortskoordinaten gleichverteilt und die Impulsverteilung ist durch die Maxwell-Verteilung mit der Temperatur als Parameter gegeben.

3.1.1 Wahrscheinlichkeitsdichte, Verteilungsfunktion, Randverteilung

Wir betrachten n Zufallsvariable x_1, x_2, \dots, x_n , die wir in einem n -Tupel

$$\vec{x} = (x_1, x_2, \dots, x_n)^T \quad (3.1)$$

zusammenfassen.

Wahrscheinlichkeitsdichte: Die Wahrscheinlichkeitsdichte $f(\vec{x})$ liefert die differentielle Wahrscheinlichkeit an einem Punkt \vec{x} :

$$dp(\vec{x}) = f(\vec{x}) dx_1 dx_2 \dots dx_n \quad (3.2)$$

Die Normierung erfolgt über den n -dimensionalen Raum Ω in dem f definiert oder ungleich Null ist:

$$\int_{\Omega} f(\vec{x}) dx_1 dx_2 \dots dx_n = 1 \quad (3.3)$$

Verteilungsfunktion: Die Verteilungsfunktion ergibt sich analog zum eindimensionalen Fall:

$$F(\vec{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(\vec{\xi}) d\xi_1 d\xi_2 \dots d\xi_n = 1 \quad (3.4)$$

Umgekehrt lässt sich die Wahrscheinlichkeitsdichte aus der Verteilungsfunktion ableiten:

$$f(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F(\vec{x}). \quad (3.5)$$

Randverteilung: Die Randverteilung einer Variablen x_i ist die Projektion der Wahrscheinlichkeit auf die i -te Koordinate, das heißt man betrachtet die Verteilung von x_i gemittelt über alle anderen Variablen. Zum Beispiel ist die Randverteilung von x_1 :

$$h_1(x_1) = \int_{-\infty}^{+\infty} dx_2 \int_{-\infty}^{+\infty} dx_3 \dots \int_{-\infty}^{+\infty} dx_n f(\vec{x}) \quad (3.6)$$

Beispiel: Die Aufenthaltswahrscheinlichkeit des Elektrons in einem Wasserstoffatom wird in der Regel durch Kugelkoordinaten (r, θ, ϕ) angegeben. Wenn man nur an der radialen Abhängigkeit interessiert ist, erhält man die Randverteilung von r :

$$\rho_r(r) = \int_{-1}^{+1} d \cos \theta \int_0^{2\pi} d\phi \rho(r, \theta, \phi) \quad (3.7)$$

3.1.2 Bedingte Wahrscheinlichkeitsdichten, Selektionsschnitte

Häufig möchte man Wahrscheinlichkeitsdichten betrachten unter der Bedingung, dass eine der Variablen einen bestimmten Wert hat, zum Beispiel $x_1 = x_{10}$ (Abb. 3.1a):

$$f^*(x_2, x_3, \dots, x_n | x_1 = x_{10}) = \frac{f(x_1 = x_{10}, x_2, \dots, x_n)}{h_1(x_1 = x_{10})} \quad (3.8)$$

Das entspricht einer Umnormierung der Wahrscheinlichkeitsdichte auf eine $n-1$ -dimensionale Hyperfläche, die durch $x_1 = x_{10}$ festgelegt ist.

Tatsächlich gibt man in der Praxis meistens ein endliches Intervall $x_{1L} < x_1 < x_{1H}$ vor und die Wahrscheinlichkeitsdichte für x_2, x_3, \dots, x_n muss auf diesen beschränkten n -dimensionalen Unterraum umnormiert werden (Abb. 3.1b):

$$f^*(x_2, x_3, \dots, x_n | x_{1L} < x_1 < x_{1H}) = \frac{\int_{x_{1L}}^{x_{1H}} f(x_1, x_2, \dots, x_n) dx_1}{\int_{x_{1L}}^{x_{1H}} h_1(x_1) dx_1} \quad (3.9)$$

Solche Einschränkungen von Variablenbereichen ist bei multi-dimensionalen Datensätzen ein Standardverfahren zur Bereinigung der Daten von Untergrund und

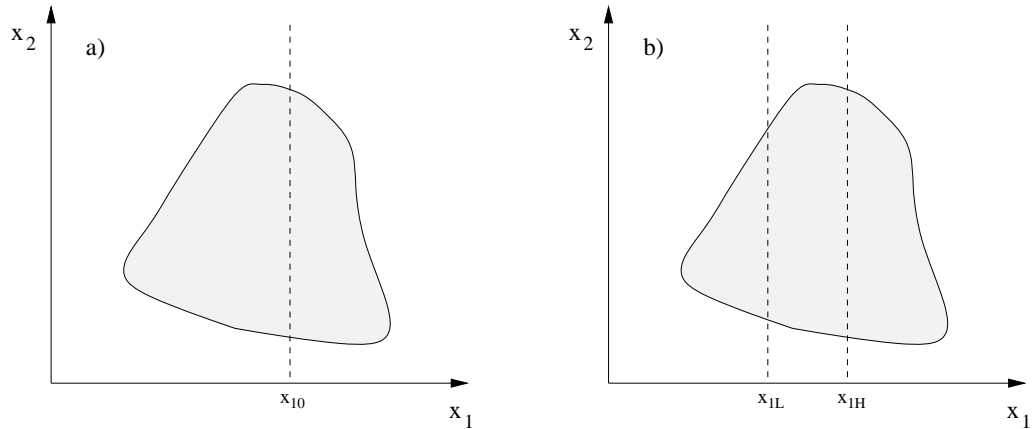


Abbildung 3.1: Bedingte Wahrscheinlichkeiten: a) Definition einer 'Hyperebene' durch $x_1 = x_{10}$, b) Schnitt in der Variablen x_1 .

zur Untersuchung von Abhängigkeiten der Variablen untereinander. Häufig versucht man Signale, die auf einem Untergrund sitzen, dadurch statistisch signifikanter zu machen, indem man Bereiche, die einen relativ hohen Untergrundbeitrag liefern wegschneidet (Selektionsschnitte).

3.2 Erwartungswerte

Erwartungswert und Varianz einer Funktion: Der Erwartungswert einer Funktion g der Zufallsvariablen $\vec{x} = (x_1, x_2, \dots, x_n)$, die die Wahrscheinlichkeitsdichte $f(\vec{x})$ haben, ist analog zum eindimensionalen Fall definiert:

$$E(g(\vec{x})) = \langle g(\vec{x}) \rangle = \int_{\Omega} g(\vec{x}) f(\vec{x}) dx_1 dx_2 \dots dx_n \quad (3.10)$$

Entsprechend ist die Varianz der Funktion g :

$$V(g(\vec{x})) = E((g(\vec{x}) - E(g(\vec{x})))^2) = \int_{\Omega} (g(\vec{x}) - \langle g(\vec{x}) \rangle)^2 f(\vec{x}) dx_1 dx_2 \dots dx_n \quad (3.11)$$

Momente: In Erweiterung der Definition für die Momente einer eindimensionalen Verteilung in Abschnitt 1.2.2 werden Momente einer mehrdimensionalen Verteilung als Erwartungswerte von Produkten von Potenzen der Zufallszahlen definiert:

1. Momente um den Ursprung:

$$\lambda_{l_1 l_2 \dots l_n} = E(x_1^{l_1} \cdot x_2^{l_2} \cdot \dots \cdot x_n^{l_n}) \quad (3.12)$$

2. Zentrale Momente:

$$\mu_{l_1 l_2 \dots l_n} = E((x_1 - \mu_1)^{l_1} \cdot (x_2 - \mu_2)^{l_2} \cdot \dots \cdot (x_n - \mu_n)^{l_n}) \quad (3.13)$$

Dabei sind die niedrigsten Momente die Mittelwerte μ_i der Zufallsvariablen x_i , die den niedrigsten Momenten mit $l_i = 1, l_k = 0$ für $k \neq i$ entsprechen:

$$\mu_i = \int_{\Omega} x_i f(\vec{x}) dx_1 dx_2 \dots dx_n \quad (3.14)$$

3.3 Kovarianzmatrix

3.3.1 Definition und Eigenschaften der Kovarianzmatrix

Die Momente mit $l_i = l_j = 1$; $l_k = 0$ für $k \neq i$, $k \neq j$ oder $l_i = 2$; $l_k = 0$ für $i = j$ und $k \neq i$ werden in einer sogenannten Kovarianzmatrix V_{ij} zusammengefasst:

$$V_{ij} = \mu_{0 \dots \underbrace{1}_i \dots \underbrace{1}_j \dots 0} = E((x_i - \mu_i)(x_j - \mu_j)) \quad (3.15)$$

$$V_{ii} = \mu_{0 \dots \underbrace{2}_i \dots 0 \dots 0} = E((x_i - \mu_i)^2) \quad (3.16)$$

Die Kovarianzmatrix hat folgende Eigenschaften:

1. Die Matrix ist symmetrisch:

$$V_{ij} = V_{ji}. \quad (3.17)$$

2. Für $i = j$ ergibt sich die Varianz von x_i :

$$V_{ii} = E((x_i - \mu_i)^2) = E(x_i^2) - (E(x_i))^2 = \sigma_i^2 \geq 0. \quad (3.18)$$

3. Die nicht-diagonalen Elemente, $i \neq j$, sind die Kovarianzen:

$$V_{ij} = \text{cov}(x_i, x_j) = E((x_i - \mu_i)(x_j - \mu_j)) = E(x_i x_j) - E(x_i) E(x_j) \stackrel{\leq}{\geq} 0. \quad (3.19)$$

3.3.2 Beispiel: Multi-dimensionale Gaussverteilung

Durch Verallgemeinerung der Varianz σ^2 auf die Kovarianzmatrix wird eine mehrdimensionale Gauss- oder Normalverteilung definiert:

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n \det(V)}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})\right) \quad (3.20)$$

Bei zwei Variablen x_1, x_2 ist die Kovarianzmatrix:

$$V = \begin{pmatrix} \sigma_1^2 & \text{cov}(x_1, x_2) \\ \text{cov}(x_1, x_2) & \sigma_2^2 \end{pmatrix} \quad (3.21)$$

Die inverse Kovarianzmatrix ist:

$$V^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - (\text{cov}(x_1, x_2))^2} \begin{pmatrix} \sigma_2^2 & -\text{cov}(x_1, x_2) \\ -\text{cov}(x_1, x_2) & \sigma_1^2 \end{pmatrix} \quad (3.22)$$

Für einen festen Wert des Exponenten in (3.20) beschreibt $f(x)$ eine Kontur mit fester Wahrscheinlichkeitsdichte

$$f_{\text{Kontur}} = f(x | (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu}) = \text{const}). \quad (3.23)$$

Im Falle der multi-dimensionalen Gauss-Verteilung sind die Konturen konstanter Wahrscheinlichkeitsdichte n-dimensionale Ellipsoide.

Wenn die Kovarianzmatrix und damit auch ihre inverse Matrix diagonal sind, folgt für den Exponenten der Gauss-Verteilung (3.20):

$$(\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu}) = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad (3.24)$$

Es treten also keine gemischten Terme $x_i \cdot x_j$ mit $i \neq j$ auf. Deshalb lässt sich in diesem Fall die mehrdimensionale Gauss-Verteilung (3.20) in ein Produkt eindimensionaler Gauss-Verteilungen zerlegen:

$$f(\vec{x}) = \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \quad (3.25)$$

Da V und V^{-1} symmetrische, positiv definite Matrizen sind, lässt sich immer eine orthogonale Transformation $x_i \rightarrow x'_i$ finden, so dass V' und V'^{-1} diagonal sind (Hauptachsentransformation):

$$\vec{x}^T V^{-1} \vec{x} = \vec{x}'^T U^{-1} U V^{-1} U^{-1} U \vec{x} \quad (3.26)$$

Für orthogonale Transformationen gilt $U^T = U^{-1}$. Die Transformation U wird so bestimmt, dass $U V^{-1} U^{-1}$ diagonal ist.

Häufig sind auf Computersystemen bereits Generatoren für gauss-verteilte Zufallszahlen vorhanden. Um mehrdimensionale Gauss-Verteilungen zu erzeugen, bestimmt man zunächst die Transformation U , die V^{-1} diagonal macht. Die Diagonalelemente $\sigma_i'^2$ und die transformierten Mittelwerte $\mu'_i = U_{ij} \mu_j$ sind die Parameter von n unabhängigen Gauss-Verteilungen. Entsprechend diesen Verteilungen erzeugt man nun n unabhängige gauss-verteilte Zufallszahlen x'_i , die dann mittels $x_i = U_{ij}^{-1} x'_j = U_{ji} x'_j$ zurücktransformiert werden.

3.3.3 Kovarianzen von Stichproben

In Analogie zu der Schätzung der Varianz aus einer Stichprobe in (1.52) werden die Kovarianzen geschätzt. Die Korrelation zwischen zwei Variablen x^j, x^k , deren Verteilung an den Messpunkten i abgetastet wird, ergeben sich zu:

$$\text{cov}(x^j, x^k) = \frac{1}{n-1} \sum_{i=1}^n (x_i^j - \bar{x}^j)(x_i^k - \bar{x}^k) \quad (3.27)$$

3.3.4 Kovarianzmatrix von unabhängigen Variablen

Wenn die Zufallsvariablen x_i unabhängig sind, faktorisiert die Wahrscheinlichkeitsdichte:

$$f(\vec{x}) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \quad (3.28)$$

Wie bei der Gauss-Verteilung (3.25) ist auch im allgemeinen Fall die Kovarianzmatrix von unabhängigen Variablen diagonal. Um die Kovarianzmatrix auszurechnen, berechnen wir zunächst den Erwartungswert von $x_i x_j$:

$$E(x_i x_j) = \int x_i f_i(x_i) dx_i \cdot \int x_j f_j(x_j) dx_j \cdot \prod_{k \neq i, k \neq j} \underbrace{\int f_k(x_k) dx_k}_{=1} = E(x_i) \cdot E(x_j) \quad (3.29)$$

Damit ergibt sich:

$$\text{cov}(x_i, x_j) = E((x_i - \mu_i)(x_j - \mu_j)) = \underbrace{E(x_i x_j) - E(x_i) E(x_j)}_{(3.29)} = 0 \quad (3.30)$$

Für unabhängige Variable verschwinden also die Kovarianzen:

$$x_i, x_j \text{ unabhängig} \implies \text{cov}(x_i, x_j) = 0 \quad (3.31)$$

Die Umkehrung dieses Satzes gilt nicht im Allgemeinen. Man sieht an (3.30), dass die Kovarianzen verschwinden, wenn sich die Terme $(x_i - \mu_i)(x_j - \mu_j)$ im Mittel auslöschen. Das kann auf verschiedenste Weisen passieren. Zum Beispiel heben sich in Abb. 3.2b gerade die Kovarianzen der rechten und linken Hälfte der Verteilung auf (in der linken Hälfte ergibt sich eine positive Korrelation und in der rechten eine negative). Die Kovarianz der gesamten Verteilung verschwindet also, obwohl es offensichtlich eine Abhängigkeit von x_1 und x_2 gibt.

3.3.5 Korrelationen

Wenn die Kovarianzen nicht verschwinden, nennt man die entsprechenden Variablen korreliert. Als Maß für die Stärke der Korrelation definiert man den **Korrelationskoeffizienten**:

$$\rho(x_i, x_j) = \frac{V_{ij}}{\sqrt{V_{ii} V_{jj}}} = \frac{\text{cov}(x_i, x_j)}{\sigma_i \cdot \sigma_j} \quad (3.32)$$

Durch die Normierung auf die Standardabweichungen ergibt sich für den Wertebereich von ρ :

$$-1 \leq \rho(x_i, x_j) \leq +1 \quad (3.33)$$

Je mehr der Korrelationskoeffizient von Null abweicht, umso besser kann man aus der Kenntnis einer Variablen die andere vorhersagen (Abb. 3.2):

$$\begin{aligned} \rho(x_i, x_j) \rightarrow +1 &\implies x_i \rightarrow +x_j && \text{(positiv korreliert)} \\ \rho(x_i, x_j) \rightarrow \pm 0 &\implies x_i, x_j \text{ unabhängig} && \text{(nicht korreliert)} \\ \rho(x_i, x_j) \rightarrow -1 &\implies x_i \rightarrow -x_j && \text{(negativ korreliert)} \end{aligned} \quad (3.34)$$

Beispiele:

1. Ein Teilchen, das wie Abb. 3.3 durch eine Materieschicht geht, wird unter einem Winkel θ gestreut und erfährt eine Ablage Δx . Streuwinkel und Ablage sind positiv korreliert.
2. Ein Anthropologe untersucht 5 Funde von Neandertalerknochen. Er vergleicht die Längen der Oberarm- mit der der Oberschenkelknochen und möchte seinen naheliegenden Verdacht, dass beide korreliert sind, statistisch erhärten.

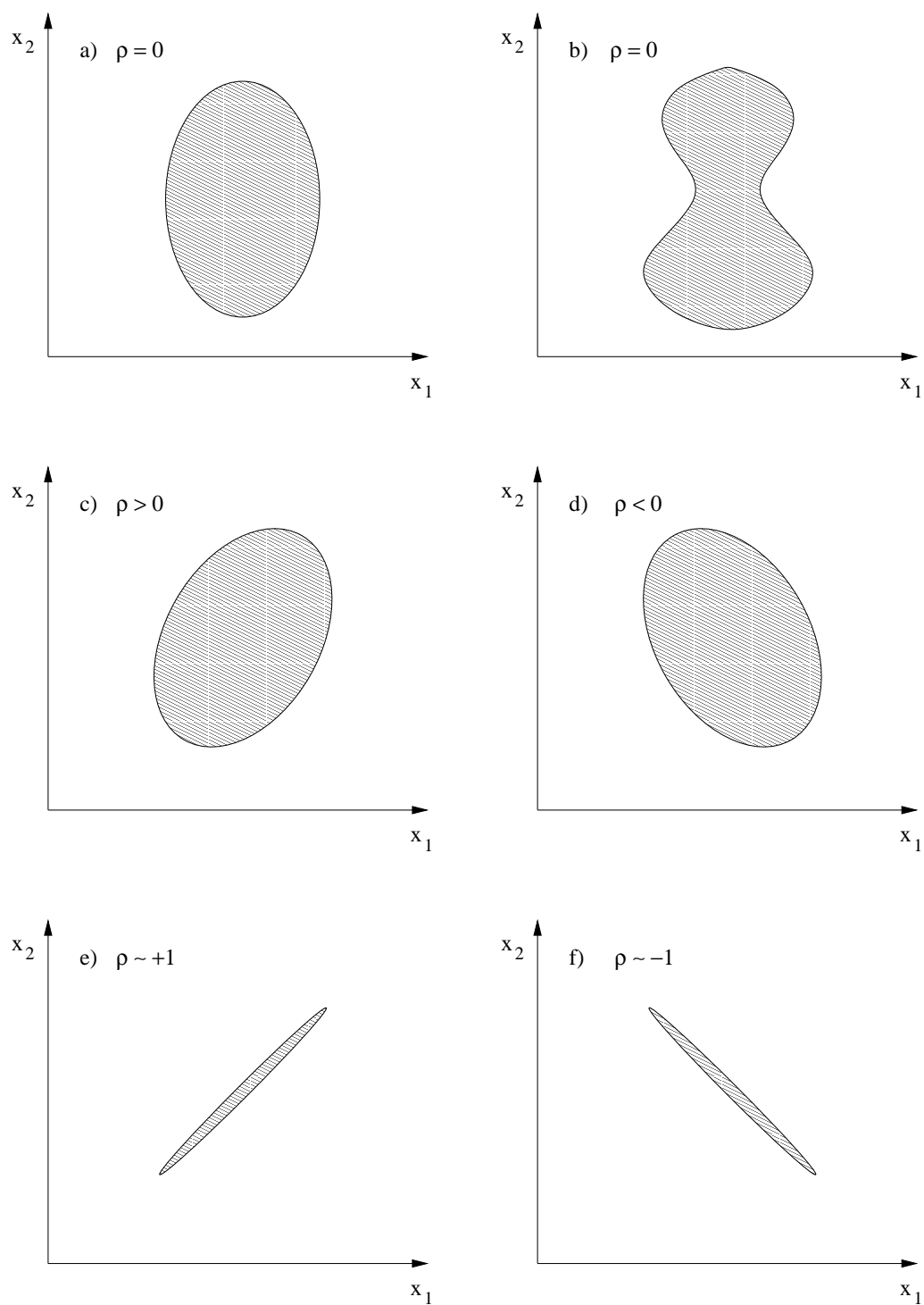


Abbildung 3.2: Verteilungsformen mit unterschiedlichem Korrelationskoeffizienten ρ .

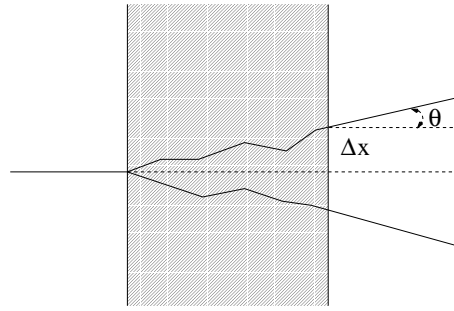


Abbildung 3.3: Streuung von Teilchen in einer Materieschicht, zum Beispiel α -Teilchen in einer Goldfolie wie bei dem Rutherford-Experiment.

Die vorliegenden Daten sind (l^a , l^b sind die Längen jeweils der Arm- und Bein-knochen):

Fund	l^a [mm]	l^b [mm]	l^{a2} [mm ²]	l^{b2} [mm ²]	$l^a l^b$ [mm ²]
1	312	430	97344	184900	134160
2	335	458	112225	209764	153430
3	286	407	81796	165649	116402
4	312	440	97344	193600	137280
5	305	422	93025	178084	128710
Mittel	310.0	431.4	96346.8	186399.4	133996.4
$\sigma_{l^{a,b}}$	17.56	19.15		$\text{cov}(l^a, l^b)$	328.0

(3.35)

Die letzten drei Spalten enthalten die Berechnung von l^{a2} , l^{b2} und $l^a \cdot l^b$ und deren Mittelwerte, die dann in die Berechnung der Kovarianzmatrix eingehen. Entsprechend (3.27) ergibt sich:

$$\text{cov}(l^a, l^b) = E(l^a \cdot l^b) - E(l^a) E(l^b) = \frac{5}{5-1} (\overline{l^a l^b} - \bar{l}^a \cdot \bar{l}^b) \quad (3.36)$$

Der Faktor $5/4$ korrigiert wie bei der Berechnung der Varianz einer Stichprobe darauf, dass bezüglich des Mittelwertes bereits die quadratischen Abweichungen minimiert werden. Einsetzen der Zahlen aus der Tabelle ergibt:

$$\text{cov}(l^a, l^b) = 328.0 \implies \rho(l^a, l^b) = \frac{\text{cov}(l^a, l^b)}{\sigma_{l^a} \cdot \sigma_{l^b}} = 0.975 \quad (3.37)$$

Die Korrelation in der Größe der Arm- und Bein-knochen ist also sehr hoch.

3.4 Lineare Funktionen von mehreren Zufallsvariablen

In den folgenden Abschnitten werden Funktionen von mehreren Zufallsvariablen betrachtet. Wir interessieren uns insbesondere für die Berechnung einfacher Erwartungswerte dieser Funktionen, wie Mittelwerte und Varianzen. Die Berechnung der

Varianz einer Funktion von Zufallsvariablen wird für die Fehlerfortpflanzung von Messungen benutzt.

Ein besonders einfacher Fall ist eine lineare Funktion von mehreren Variablen. Wir werden im folgenden häufig auch bei nicht-linearen Funktionen durch Linearisierung um einen Entwicklungspunkt die Ergebnisse für lineare Funktionen benutzen.

Es sei g eine lineare Funktion der n Zufallsvariablen $\vec{x} = (x_1, \dots, x_n)$:

$$g(\vec{x}) = \sum_{i=1}^n a_i x_i \quad (3.38)$$

Erwartungswert: Der Erwartungswert der Funktion ist:

$$E(g(\vec{x})) = \sum_{i=1}^n a_i \underbrace{E(x_i)}_{=\mu_i} = \sum_{i=1}^n a_i \mu_i \quad (3.39)$$

Varianz:

$$\begin{aligned} V(g(\vec{x})) &= E((g(\vec{x}) - E(g(\vec{x})))^2) = E((\sum_i a_i x_i - \sum_i a_i \mu_i)^2) \\ &= E((\sum_i a_i (x_i - \mu_i))^2) = \sum_i \sum_j a_i a_j E((x_i - \mu_i)(x_j - \mu_j)) \\ &= \sum_i \sum_j a_i a_j V_{ij} \end{aligned} \quad (3.40)$$

Dabei ist V_{ij} die Kovarianzmatrix der Zufallsvariablen \vec{x} . Mit der Beziehung $V_{ij} = V_{ji}$ lässt sich die Varianz von g durch die Varianzen und die Kovarianzen ausdrücken:

$$V(g(\vec{x})) = \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j V_{ij} \quad (3.41)$$

Wenn die x_i unabhängig sind, ist $V_{ij} = 0$ für $i \neq j$ und die Varianz von g ergibt sich nur aus den Varianzen der x_i :

$$V(g(\vec{x})) = \sum_{i=1}^n a_i^2 \sigma_i^2 \quad (3.42)$$

Beispiele:

1. Eine Stichprobe x_1, \dots, x_n aus einer Verteilung mit dem Mittelwert μ und Varianz σ^2 kann man als einen Satz von n unabhängigen Zufallsvariablen interpretieren, die alle den gleichen Mittelwert $\mu_i = \mu$ und die gleiche Varianz $\sigma_i^2 = \sigma^2$ haben. Das arithmetische Mittel der x_i ist eine lineare Funktion der x_i :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.43)$$

Der Erwartungswert des Mittelwertes ist dann:

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \cdot n \mu = \mu \quad (3.44)$$

Das heisst, das arithmetischen Mittel einer Stichprobe ist eine ‘erwartungstreue’ Schätzung des Erwartungswertes μ der entsprechenden Verteilung, aus der die Stichprobe gezogen wurde.

Die Varianz des arithmetischen Mittels ist (die Kovarianzen fallen weg, weil die x_i unabhängig sind):

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \left(\frac{1}{n}\right)^2 \sum_i \sigma_i^2 = \left(\frac{1}{n}\right)^2 n \sigma^2 = \frac{\sigma^2}{n} \quad (3.45)$$

Damit hat man das bekannte Ergebnis, dass der Fehler des Mittelwertes von n Messungen um $1/\sqrt{n}$ kleiner als der Fehler der Einzelmessung ist:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (3.46)$$

2. Im allgemeinen hat die Varianz einer Funktion von zwei Zufallsvariablen,

$$g(x, y) = a x + b y, \quad (3.47)$$

folgende Form:

$$V(ax + by) = a^2 \underbrace{V_{xx}}_{=\sigma_x^2} + b^2 \underbrace{V_{yy}}_{=\sigma_y^2} + 2ab \underbrace{V_{xy}}_{=\text{cov}(x,y)} = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab \sigma_x \sigma_y \rho(x, y) \quad (3.48)$$

Dabei kann der Korrelationskoeffizient $\rho(x, y)$ Werte von -1 bis +1 annehmen.

3.5 Nicht-lineare Funktionen von Zufallsvariablen

3.5.1 Eine Funktion von einem Satz von Zufallsvariablen

In diesem Abschnitt wollen wir allgemeine Funktionen g der Zufallsvariablen betrachten:

$$g = g(x_1, \dots, x_n). \quad (3.49)$$

Um die Ergebnisse des vorigen Abschnitts benutzen zu können, linearisieren wir die Funktion in der Umgebung der Mittelwerte $\vec{\mu}$:

$$g(\vec{x}) = g(\vec{\mu}) + \sum_{i=1}^n (x_i - \mu_i) \frac{\partial g}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} + \dots \quad (3.50)$$

Erwartungswert: Der Erwartungswert der Funktion g ist in der linearen Näherung:

$$E(g(\vec{x})) = E(g(\vec{\mu})) + \sum_{i=1}^n \underbrace{E(x_i - \mu_i)}_{=0} \frac{\partial g}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} = E(g(\vec{\mu})) = g(\vec{\mu}) \quad (3.51)$$

Der Erwartungswert der Funktion $g(\vec{x})$ ist also diese Funktion an der Stelle der Erwartungswerte von \vec{x} :

$$E(g(\vec{x})) = g(\vec{\mu}) \quad (3.52)$$

Varianz:

$$\begin{aligned}
 V(g(\vec{x})) &= E((g(\vec{x}) - E(g(\vec{x})))^2) \\
 &= E((g(\vec{x}) - g(\vec{\mu}))^2) \\
 &= E\left(\left(\sum_i (x_i - \mu_i) \frac{\partial g}{\partial x_i}\right)^2\right) \\
 &= \sum_i \sum_j \frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} E((x_i - \mu_i)(x_j - \mu_j)) \\
 &= \sum_i \sum_j \frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} V_{ij}
 \end{aligned} \tag{3.53}$$

Das entspricht also genau dem Ergebnis (3.40), wenn man statt der Koeffizienten a_i die partiellen Ableitungen $\partial g/\partial x_i$ einsetzt.

In Matrixschreibweise definiert man den Spaltenvektor:

$$\vec{a} = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix} \tag{3.54}$$

Damit ergibt sich für die Varianz:

$$V(g(\vec{x})) = \sigma^2(g(\vec{x})) = \vec{a}^T V(\vec{x}) \vec{a} \tag{3.55}$$

Zum Beispiel erhält man für $n = 2$:

$$\sigma^2(g(\vec{x})) = \left(\frac{\partial g}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial x_2}\right)^2 \sigma_2^2 + 2 \frac{\partial g}{\partial x_1} \frac{\partial g}{\partial x_2} \text{cov}(x_1, x_2) \tag{3.56}$$

Das ist also die bekannte Formel, die auch für Fehlerfortpflanzung benutzt wird.

3.5.2 Mehrere Funktionen von einem Satz von Zufallszahlen

Wir betrachten jetzt den allgemeineren Fall, dass m Funktionen $g = (g_1, \dots, g_m)$ von den gleichen n Zufallszahlen (x_1, \dots, x_n) abhängen:

$$\vec{g}(\vec{x}) = \begin{pmatrix} g_1(\vec{x}) \\ \vdots \\ g_m(\vec{x}) \end{pmatrix} \tag{3.57}$$

Ein häufig auftretendes Beispiel ist eine Koordinatentransformation der Zufallsvariablen: die transformierten Variablen sind im allgemeinen eine Funktion aller ursprünglichen Variablen.

Die Erwartungswerte der Funktionen g_j und deren Varianzen ergeben sich für jede Funktion einzeln. Neu kommt jetzt allerdings hinzu, dass die Funktionen untereinander korreliert sein können und damit nicht-verschwindende Kovarianzen haben.

Wir linearisieren wieder jede der Funktionen ($k = 1, \dots, m$):

$$g_k(\vec{x}) = g_k(\vec{\mu}) + \sum_{i=1}^n (x_i - \mu_i) \frac{\partial g_k}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} + \dots \tag{3.58}$$

Mit

$$S_{ki} = \left. \frac{\partial g_k}{\partial x_i} \right|_{\vec{x}=\vec{\mu}} \quad (3.59)$$

ergibt (3.58):

$$g_k(\vec{x}) = g_k(\vec{\mu}) + \sum_{i=1}^n (x_i - \mu_i) S_{ki}$$

oder

$$\vec{g}(\vec{x}) = \vec{g}(\vec{\mu}) + S \vec{x} \quad (3.60)$$

Dabei sind \vec{x} , $\vec{\mu}$ Spaltenvektoren und die Jacobische Funktionalmatrix S ist in Matrixschreibweise:

$$S = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \quad (3.61)$$

Erwartungswert: Die Erwartungswerte der Funktionen $\vec{g}(\vec{x})$ ergibt sich wie für eine einzelne Funktion (3.51):

$$E(\vec{g}(\vec{x})) = \vec{g}(\vec{\mu}) \quad (3.62)$$

Varianz:

$$\begin{aligned} V_{kl}(\vec{g}(\vec{x})) &= E[(g_k(\vec{x}) - E[g_k(\vec{x})])(g_l(\vec{x}) - E[g_l(\vec{x})])] \\ &= \sum_i \sum_j \frac{\partial g_k}{\partial x_i} \frac{\partial g_l}{\partial x_j} \underbrace{E((x_i - \mu_i)(x_j - \mu_j))}_{=V_{ij}(\vec{x})} \\ &= \sum_i \sum_j \frac{\partial g_k}{\partial x_i} \frac{\partial g_l}{\partial x_j} V_{ij}(\vec{x}) = \sum_i \sum_j S_{ki} S_{lj} V_{ij}(\vec{x}) \end{aligned} \quad (3.63)$$

$$\implies V(\vec{g}(\vec{x})) = S \cdot V(\vec{x}) \cdot S^T$$

Dabei sind in der letzten Zeile alle Größen Matrizen.

Um das obige Beispiel einer Variablentransformation aufzugreifen: Die Matrix S kann man beispielsweise so bestimmen, dass die Transformation $\vec{x} \rightarrow \vec{g}$ die Kovarianzmatrix $V(\vec{g})$ diagonal macht, die neuen Variablen g_i also nicht korreliert sind.

Beispiel: Fehlerfortpflanzung bei Koordinatenwechsel.

Auf einem Koordinatenmesstisch werden rechtwinklige Koordinaten (x, y) mit den Auflösungen

$$\begin{aligned} \sigma_x &= 1 \mu\text{m} \\ \sigma_y &= 3 \mu\text{m} \end{aligned} \quad (3.64)$$

gemessen. Da die Messungen der beiden Koordinaten unabhängig sein sollen, ist die Kovarianzmatrix diagonal:

$$V(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix} \quad (3.65)$$

Für die weitere Auswertung sollen die Messpunkte in Polarkoordinaten (r, ϕ) ausgedrückt werden:

$$\begin{aligned} x &= r \cos \phi \\ y &= r \sin \phi \end{aligned} \quad \implies \quad \begin{aligned} r &= \sqrt{x^2 + y^2} \\ \phi &= \arctan \frac{y}{x} \end{aligned} \quad (3.66)$$

Wir wollen nun berechnen, wie sich der Fehler der x, y -Messungen auf r, ϕ fort-pflanzt und bestimmen deshalb die Kovarianzmatrix für die Variablen r, ϕ . Die Funktionalmatrix für die Transformation ist:

$$\begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{pmatrix} \quad (3.67)$$

Damit transformiert sich die Kovarianzmatrix wie folgt:

$$V(r, \phi) = S \cdot V(x, y) \cdot S^T = \begin{pmatrix} \frac{1}{r^2}(x^2\sigma_x^2 + y^2\sigma_y^2) & \frac{xy}{r^3}(-\sigma_x^2 + \sigma_y^2) \\ \frac{xy}{r^3}(-\sigma_x^2 + \sigma_y^2) & \frac{1}{r^4}(y^2\sigma_x^2 + x^2\sigma_y^2) \end{pmatrix} \quad (3.68)$$

Ausgedrückt in Polarkoordinaten ergibt sich für die Kovarianzmatrix:

$$\begin{aligned} V(r, \phi) &= \begin{pmatrix} \sigma_r^2 & \text{cov}(r, \phi) \\ \text{cov}(r, \phi) & \sigma_\phi^2 \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\phi \sigma_x^2 + \sin^2\phi \sigma_y^2 & \frac{\sin\phi \cos\phi}{r}(-\sigma_x^2 + \sigma_y^2) \\ \frac{\sin\phi \cos\phi}{r}(-\sigma_x^2 + \sigma_y^2) & \frac{1}{r^2}(\sin^2\phi \sigma_x^2 + \cos^2\phi \sigma_y^2) \end{pmatrix} \end{aligned} \quad (3.69)$$

Man sieht, dass die Kovarianzmatrix auch in Polarkoordinaten diagonal ist, wenn die x - und y -Messgenauigkeit gleich, also $\sigma_x = \sigma_y$, ist. Die Kovarianzen verschwinden auch für die Spezialfälle $\phi = 0^\circ, 90^\circ$, das heisst für Punkte auf der x - bzw. y -Achse:

$$V(r, \phi = 0^\circ) = \begin{pmatrix} \sigma_r^2 = \sigma_x^2 & \text{cov}(r, \phi) = 0 \\ \text{cov}(r, \phi) = 0 & \sigma_\phi^2 = \sigma_y^2 \end{pmatrix} \quad (3.70)$$

$$V(r, \phi = 90^\circ) = \begin{pmatrix} \sigma_r^2 = \sigma_y^2 & \text{cov}(r, \phi) = 0 \\ \text{cov}(r, \phi) = 0 & \sigma_\phi^2 = \sigma_x^2 \end{pmatrix} \quad (3.71)$$

Man kann jetzt auch wieder umgekehrt die Varianzen der Zufallsvariablen x und y berechnen, wenn die Kovarianzmatrix in Polarkoordinaten vorliegt. Will man zum Beispiel die Varianz von x am Punkt $(1, 1)$, also $(r = \sqrt{2}, \phi = 45^\circ)$, berechnet man zunächst

$$\sigma_r^2 = 5, \quad \sigma_\phi = \frac{5}{2}, \quad \text{cov}(r, \phi) = \frac{4}{\sqrt{2}} \quad (3.72)$$

Damit ergibt sich beispielsweise für σ_x^2 (siehe (3.56)):

$$\begin{aligned} \sigma_x^2 &= \left(\frac{\partial x}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial x}{\partial \phi}\right)^2 \sigma_\phi^2 + 2 \frac{\partial x}{\partial r} \frac{\partial x}{\partial \phi} \text{cov}(r, \phi) \\ &= \cos^2\phi \sigma_r^2 + r^2 \sin^2\phi \sigma_\phi^2 - 2r \cos\phi \sin\phi \text{cov}(r, \phi) \\ &= \frac{5}{2} + \frac{5}{2} - \frac{8}{2} = 1 \quad (= \sigma_x^2) \end{aligned} \quad (3.73)$$

Es ergibt sich also korrekt wieder der Wert $\sigma_x^2 = 1$, der hineingesteckt wurde. Hier sieht man, dass man im allgemeinen die Kovarianzen nicht vernachlässigen kann: ohne Berücksichtigung der Kovarianz hätte sich $\sigma_x^2 = 5$ ergeben.

3.6 Transformationen von Zufallsvariablen

In dem obigen Beispiel hatten wir eine Transformation der Zufallsvariablen x, y auf r, ϕ und die daraus folgende Transformation der Varianzen betrachtet. Wir fragen nun, wie sich die Wahrscheinlichkeitsdichten transformieren, wenn man zu anderen Variablen übergeht. Variablentransformationen macht man unter anderem auch um einfachere Wahrscheinlichkeitsdichten zu erhalten, zum Beispiel Gleichverteilungen für eine Simulation (siehe Abschnitt 1.3).

Wir betrachten zunächst den Fall, dass eine einzelne Variable in eine andere transformiert wird:

$$x \rightarrow z, \quad f(x) \rightarrow g(z) \quad (3.74)$$

In einem Intervall dx , das in dz übergeht, müssen die Wahrscheinlichkeiten vor und nach der Transformation gleich sein:

$$dp = f(x) dx = g(z) dz \implies g(z) = f(x(z)) \left| \frac{dx}{dz} \right| \quad (3.75)$$

Im rechten Ausdruck wird der Betrag der Ableitung genommen, damit die Wahrscheinlichkeit positiv bleibt.

Für n Variable mit der Transformation

$$(x_1, \dots, x_n) \rightarrow (z_1, \dots, z_n), \quad f(x_1, \dots, x_n) \rightarrow g(z_1, \dots, z_n) \quad (3.76)$$

ergibt sich die Bedingung:

$$f(\vec{x}) dx_1 \dots dx_n = g(\vec{z}) dz_1 \dots dz_n \implies g(\vec{z}) = f(\vec{x}(\vec{z})) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(z_1, \dots, z_n)} \right| \quad (3.77)$$

Der rechte Ausdruck ist die Funktional- oder Jacobi-Determinante:

$$\left| \frac{\partial(x_1, \dots, x_n)}{\partial(z_1, \dots, z_n)} \right| = \det \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} & \dots & \frac{\partial x_1}{\partial z_n} \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} & \dots & \frac{\partial x_2}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \frac{\partial x_n}{\partial z_2} & \dots & \frac{\partial x_n}{\partial z_n} \end{pmatrix} \quad (3.78)$$

Beispiele:

1. In der Physik kommt häufig die Transformation auf krummlinige Koordinaten vor. Zum Beispiel transformiert sich bei dem Übergang von kartesischen auf Kugelkoordinaten, $(x, y, z) \rightarrow (r, \theta, \phi)$, das Volumenelement bekanntlich wie

$$dx dy dz \rightarrow r^2 \sin \theta dr d\theta d\phi, \quad (3.79)$$

so dass sich die Jacobi-Determinante zu $r^2 \sin \theta$ ergibt.

2. Ein schnelles geladenes Teilchen emittiert sogenannte Bremsstrahlung, wenn eine Kraft auf das Teilchen wirkt, wie beim Durchgang durch Materie oder

in elementaren Wechselwirkungen. Die Wahrscheinlichkeitsdichte für die Abstrahlungsrichtung θ relativ zur Teilchenrichtung hat etwa folgende Form:

$$w(\theta) = w_0 \frac{\sin \theta}{1 - \beta \cos \theta} \quad (3.80)$$

Dabei ist $\beta = v/c$ die Teilchengeschwindigkeit in Einheiten der Lichtgeschwindigkeit. Für Elektronen ist β schon bei relativ niedrigen Energien sehr nahe 1, zum Beispiel für $E = 1 \text{ GeV}$ ist $1 - \beta = 1.3 \cdot 10^{-7}$. In diesem Fall ‘hochrelativistischer’ Teilchen ist der Ausdruck $1/(1 - \beta \cos \theta)$ bei $\theta = 0$ nahezu divergent. Dieses Verhalten wird auch nicht durch den $\sin \theta$ -Term in (3.80) gedämpft, weil das Winkelement $\sin \theta d\theta = d \cos \theta$ bei $\theta = 0$ endlich bleibt.

Eine Simulation der Abstrahlung wird also zum Beispiel mit der ‘Hit and Miss’ Methode sehr ineffektiv. Man wird also eine Transformation suchen, die das Polverhalten dämpft. Tatsächlich kann man (3.80) auf eine Gleichverteilung transformieren. Entsprechend Abschnitt 1.3 machen wir den Ansatz (u ist eine zwischen 0 und 1 gleichverteilte Zufallsvariable):

$$w(\theta) d\theta = du \implies u = w_0 \int_0^\theta w(\vartheta) d\vartheta = W(\theta) = \frac{w_0}{\beta} \ln \frac{1 + \beta \cos \theta}{1 - \beta}, \quad (3.81)$$

wobei $W(\theta)$ die Verteilungsfunktion ist. Der Normierungsfaktor w_0 ergibt sich aus der Integration von $w(\theta)$ über den gesamten Wertebereich:

$$\frac{1}{w_0} = \int_0^\pi w(\vartheta) d\vartheta = W(\pi) = \frac{1}{\beta} \ln \frac{1 + \beta}{1 - \beta}, \quad (3.82)$$

Die Transformation $\theta \rightarrow u$ ergibt sich aus der Inversion von (3.81):

$$\theta = \arccos \left[\frac{1}{\beta} \left((1 - \beta) e^{\frac{\beta u}{w_0}} - 1 \right) \right] \quad (3.83)$$

Nehmen wir weiterhin an, dass die azimuthale Winkelverteilung der Strahlung durch Polarisierungseffekte (die Elektronenspins könnten zum Beispiel transversal zu ihrer Flugrichtung polarisiert sein) sinusförmig moduliert wird:

$$w'(\theta, \phi) = w'_0 \frac{\sin \theta \sin \phi}{1 - \beta \cos \theta} \quad (3.84)$$

Eine entsprechende Transformation von ϕ im Intervall 0 bis π auf eine zwischen 0 und 1 gleichverteilte Variable v erhält man wie in (3.81):

$$\alpha \sin \phi d\phi = dv \implies v = \frac{\int_0^\phi \sin \varphi d\varphi}{\int_0^\pi \sin \varphi d\varphi} = \frac{\cos \phi + 1}{2} \quad (3.85)$$

Dabei ist $\alpha = 1/2$ die Normierungskonstante und es gilt $w'_0 = w_0 \alpha$. Die gesamte Variablentransformation ist damit:

$$\begin{aligned} \theta &= \arccos \left[\frac{1}{\beta} \left((1 - \beta) e^{\frac{\beta u}{w_0}} - 1 \right) \right] \\ \phi &= \arccos(2v - 1) \end{aligned} \quad (3.86)$$

Daraus ergibt sich die Funktionaldeterminante:

$$\left| \frac{\partial(\theta, \phi)}{\partial(u, v)} \right| = \det \begin{pmatrix} \frac{\partial\theta}{\partial u} & \frac{\partial\theta}{\partial v} \\ \frac{\partial\phi}{\partial u} & \frac{\partial\phi}{\partial v} \end{pmatrix} = \frac{1}{w'_0} \frac{1 - \beta \cos \theta}{\sin \theta \sin \phi} = \frac{1}{w'(\theta, \phi)} \quad (3.87)$$

Es ist natürlich kein Zufall, dass die Jacobi-Determinante gerade das Reziproke der ursprünglichen Dichteverteilung ergibt, weil ja gerade auf eine Gleichverteilung transformiert werden sollte.