

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:

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

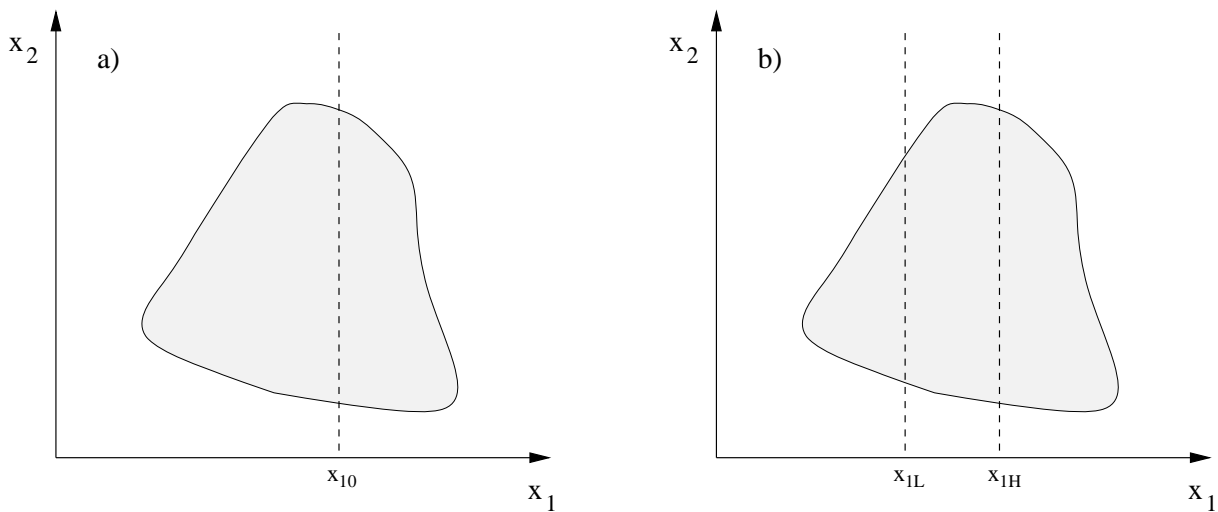


Abbildung 3.1: Bedingte Wahrscheinlichkeiten: a) Definition einer ‘Hyperebene’ durch $x_1 = x_{10}$, b) Schnitt in der Variablen x_1 .

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 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 μ_i die Mittelwerte 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:

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

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:

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

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

$$(\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu}) = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad (3.23)$$

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:

$$f(\vec{x}) = \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \quad (3.24)$$

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):

$$\vec{x}^T V^{-1} \vec{x} = \vec{x}'^T U^{-1} U V^{-1} U^{-1} U \vec{x} \quad (3.25)$$

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

Um mehrdimensionale Gauss-Verteilungen auf dem Computer 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 Kovarianzmatrix unabhängiger Variabler

Wenn die Zufallsvariablen x_i unabhängig sind, faktorisiert die Wahrscheinlichkeitsdichte:

$$f(\vec{x}) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \quad (3.26)$$

Wie bei der Gauss-Verteilung (3.24) ist auch im allgemeinen Fall die Kovarianzmatrix unabhängiger Variabler diagonal. Um die Kovarianzmatrix auszurechnen, berechnen wir zunächst den Erwartungswert von $x_i x_j$:

$$E(x_i x_j) = \int x_i f_i(x_i) dx_i \cdot \int x_j f_j(x_j) dx_j \cdot \prod_{k \neq i; k \neq j} \underbrace{\int f_k(x_k) dx_k}_{=1} = E(x_i) \cdot E(x_j) \quad (3.27)$$

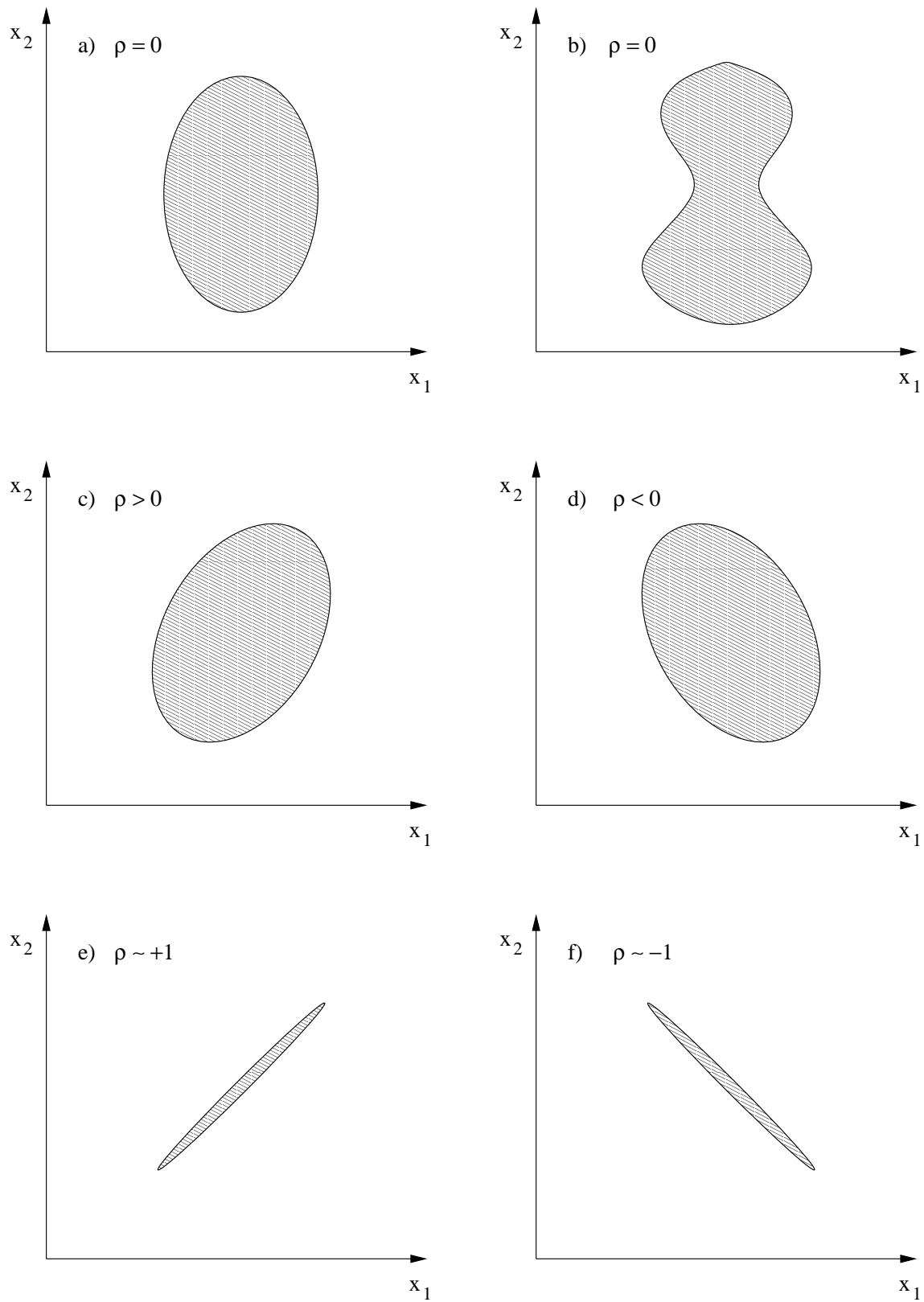
Damit ergibt sich:

$$\text{cov}(x_i, x_j) = E((x_i - \mu_i)(x_j - \mu_j)) = E(x_i x_j) - E(x_i) E(x_j) = 0 \quad (3.28)$$

Für unabhängige Variable verschwinden also die Kovarianzen:

$$x_i, x_j \text{ unabhängig} \implies \text{cov}(x_i, x_j) = 0 \quad (3.29)$$

Die Umkehrung dieses Satzes gilt nicht unbedingt. Man sieht an (3.28), dass die Kovarianzen verschwinden, wenn sich die Terme $(x_i - \mu_i)(x_j - \mu_j)$ im Mittel auslöschen. Das passiert genau dann, wenn die Vorzeichen der Abweichungen von den Mittelwerten voneinander unabhängig sind. Oder anders ausgedrückt: wenn die Verteilung der x_j um μ_j unabhängig von x_i ist und umgekehrt. In diesem Fall kann es dann trotzdem noch eine Abhängigkeit von x_i und x_j geben, indem zum Beispiel die Breiten der Verteilungen in Abhängigkeit von den anderen Variablen variieren (Abb. 3.2b).

Abbildung 3.2: Verteilungsformen mit unterschiedlichem Korrelationskoeffizienten ρ .

3.3.4 Korrelationen

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

$$\rho(x_i, x_j) = \frac{V_{ij}}{\sqrt{V_{ii} V_{jj}}} = \frac{\text{cov}(x_i, x_j)}{\sigma_i \cdot \sigma_j} \quad (3.30)$$

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

$$-1 \leq \rho(x_i, x_j) \leq +1 \quad (3.31)$$

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

$$\begin{aligned} \rho(x_i, x_j) \rightarrow +1 &\implies x_i \rightarrow +x_j && \text{(positiv korreliert)} \\ \rho(x_i, x_j) \rightarrow \pm 0 &\implies x_i, x_j \text{ unabhängig} && \text{(nicht korreliert)} \\ \rho(x_i, x_j) \rightarrow -1 &\implies x_i \rightarrow -x_j && \text{(negativ korreliert)} \end{aligned} \quad (3.32)$$

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.

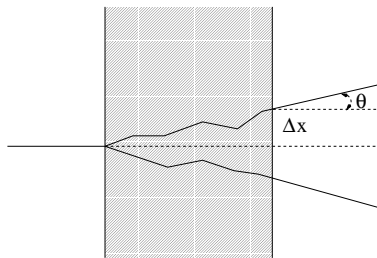


Abbildung 3.3: Streuung von Teilchen in einer Materieschicht, zum Beispiel α -Teilchen in einer Goldfolie wie bei dem Rutherford-Experiment.

2. Ein Anthropologe untersucht 5 Funde von Neanderthalerknochen. Er vergleicht die Längen der Oberarm- mit der der Oberschenkelknochen und möchte seinen naheliegenden Verdacht, dass beide korreliert sind, statistisch erhärten. Die vorliegenden Daten sind (l^a , l^b sind die Längen jeweils der Arm- und Beinknochen):

Fund	l^a [mm]	l^b [mm]	$l^a l^b$ [mm ²]
1	312	430	134160
2	335	458	153430
3	286	407	116402
4	312	440	137280
5	305	422	128710
Mittel	310.00	431.40	133996.40
σ_l	17.56	19.15	

In der letzten Spalte ist das Produkt $l^a \cdot l^b$ berechnet, dessen Mittelwert entsprechend Gleichung (3.19) gebraucht wird, um die Kovarianz zu bestimmen. Wenn wir (3.19) für unsere Stichprobe umschreiben, ergibt sich:

$$\text{cov}(l^a, l^b) = E(l^a \cdot l^b) - E(l^a) E(l^b) = \left[\left(\frac{1}{5} \sum_{i=1}^5 l_i^a \cdot l_i^b \right) - \left(\frac{1}{5} \sum_{i=1}^5 l_i^a \right) \cdot \left(\frac{1}{5} \sum_{i=1}^5 l_i^b \right) \right] \cdot \frac{5}{5-1} \quad (3.33)$$

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:

$$\text{cov}(l^a, l^b) = 262.4 \implies \rho(l^a, l^b) = \frac{\text{cov}(l^a, l^b)}{\sigma_l^a \cdot \sigma_l^b} = 0.975 \quad (3.34)$$

Die Korrelation in der Größe der Arm- und Beinknochen 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 Fehlerfortplanzung 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 Funktion der n Zufallsvariablen $\vec{x} = (x_1, \dots, x_n)$:

$$g(\vec{x}) = \sum_{i=1}^n a_i x_i \quad (3.35)$$

Erwartungswert: Der Erwartungswert der Funktion ist:

$$E(g(\vec{x})) = \sum_{i=1}^n a_i \underbrace{E(x_i)}_{=\mu_i} = \sum_{i=1}^n a_i \mu_i \quad (3.36)$$

Varianz:

$$\begin{aligned} V(g(\vec{x})) &= E((g(\vec{x}) - E(g(\vec{x})))^2) = E((\sum_i a_i x_i - \sum_i a_i \mu_i)^2) = E((\sum_i a_i (x_i - \mu_i))^2) \\ &= \sum_i \sum_j a_i a_j E((x_i - \mu_i)(x_j - \mu_j)) = \sum_i \sum_j a_i a_j V_{ij} \end{aligned} \quad (3.37)$$

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:

$$V(g(\vec{x})) = \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j V_{ij} \quad (3.38)$$

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 :

$$V(g(\vec{x})) = \sum_{i=1}^n a_i^2 \sigma_i^2 \quad (3.39)$$

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 :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.40)$$

Der Erwartungswert des Mittelwertes ist dann:

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \cdot n \mu = \mu \quad (3.41)$$

Das heisst, das arithmetische Mittel einer Stichprobe ist eine 'erwartungstreue' Schätzung von μ .

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

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \left(\frac{1}{n}\right)^2 \sum_i \sigma_i^2 = \left(\frac{1}{n}\right)^2 n \sigma^2 = \frac{\sigma^2}{n} \quad (3.42)$$

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:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (3.43)$$

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

$$g(x, y) = a x + b y, \quad (3.44)$$

folgende Form:

$$V(ax + by) = a^2 \underbrace{V_{xx}}_{=\sigma_x^2} + b^2 \underbrace{V_{yy}}_{=\sigma_y^2} + 2ab \underbrace{V_{xy}}_{=\text{cov}(x,y)} = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab \sigma_x \sigma_y \rho(x, y) \quad (3.45)$$

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:

$$g = g(x_1, \dots, x_n). \quad (3.46)$$

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

$$g(\vec{x}) = g(\vec{\mu}) + \sum_{i=1}^n (x_i - \mu_i) \frac{\partial g}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} + \dots \quad (3.47)$$

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

$$E(g(\vec{x})) = E(g(\vec{\mu})) + \sum_{i=1}^n \underbrace{E(x_i - \mu_i)}_{=0} \frac{\partial g}{\partial x_i} \Big|_{\vec{x}=\vec{\mu}} = E(g(\vec{\mu})) = g(\vec{\mu}) \quad (3.48)$$

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

$$E(g(\vec{x})) = g(\vec{\mu}) \quad (3.49)$$

Varianz:

$$\begin{aligned} V(g(\vec{x})) &= E((g(\vec{x}) - E(g(\vec{x})))^2) = E((g(\vec{x}) - g(\vec{\mu}))^2) = E\left(\left(\sum_i (x_i - \mu_i) \frac{\partial g}{\partial x_i}\right)^2\right) \\ &= \sum_i \sum_j \frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} E((x_i - \mu_i)(x_j - \mu_j)) = \sum_i \sum_j \frac{\partial g}{\partial x_i} \frac{\partial g}{\partial x_j} V_{ij} \end{aligned} \quad (3.50)$$

Das entspricht also genau dem Ergebnis (3.37), wenn man statt der Koeffizienten a_i die partiellen Ableitungen $\partial g / \partial x_i$ einsetzt.

In Matrixschreibweise definiert man den Spaltenvektor:

$$\vec{a} = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix} \quad (3.51)$$

Damit ergibt sich für die Varianz:

$$V(g(\vec{x})) = \sigma^2(g(\vec{x})) = \vec{a}^T V(\vec{x}) \vec{a} \quad (3.52)$$

Zum Beispiel erhält man für $n = 2$:

$$\sigma^2(g(\vec{x})) = \left(\frac{\partial g}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial g}{\partial x_2}\right)^2 \sigma_2^2 + 2 \frac{\partial g}{\partial x_1} \frac{\partial g}{\partial x_2} \text{cov}(x_1, x_2) \quad (3.53)$$

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