Wahrscheinlichkeitstheorie und Statistik

11. Woche

Dávid Tóth (BME SZIT)

11., 14. November 2024

Bedingter Erwartungswert

Der bedingte Erwartungswert $\mathbb{E}(Y \mid X = s)$, falls existiert, kann als eine Funktion von s betrachtet werden.

Definition. Seien X und Y Zufallsvariablen, wo X diskret und $\mathbb{E}(Y)$ endlich ist. Sei noch

$$S_X = \{s \in \mathbb{R} : \mathbb{P}(X = s) > 0\}.$$

Betrachten wir die Funktion

$$g : S_X \to \mathbb{R}, \qquad g(s) = \mathbb{E}(Y \mid X = s),$$

dann heißt g die Regressionsfunktion, und die Zufallsvariable $g(X)$ die (diskrete) Regression von Y auf X.

Notation: $\mathbb{E}(Y \mid X)$.

Bedingter Erwartungswert

Beispiel. Ein Würfel wird geworfen, und wenn die Augenzahl k ist, dann werden k Münzen geworfen. Sei X die Augenzahl, und sei Y die Anzahl der Köpfe. Was ist die Regression von Y auf X?

Dann gilt

$$\mathbb{P}(Y = k \mid X = s) = \binom{s}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{s-k}$$

für $k = 0, \ldots, s$, also

$$\mathbb{E}(Y \mid X = s) = \frac{s}{2}, \qquad \mathbb{E}(Y \mid X) = \frac{X}{2}.$$

Bedingter Erwartungswert

Erinnerung. Seien X und Y Zufallsvariablen. Die lineare Regression von Y auf X ist eine Zufallsvariable $\beta X + \alpha$ mit $\alpha, \beta \in \mathbb{R}$, für die die Quantität

$$\mathbb{E}\left((Y - (\beta X + \alpha))^2\right)$$

minimal ist.

Also ist die lineare Regression von Y auf X die "beste" Approximation von Y durch eine *lineare* Funktion von X.

Was passiert, wenn wir durch eine beliebeige Funktion approximieren können?
Was wird dann die "beste" Approximation?

Bedingter Erwartungswert

Was wäre besser als der durchschnittliche Wert von Y, wenn der Wert von X bekannt ist? Das wird gerade durch die Funktion

$$s \mapsto \mathbb{E}(Y \mid X = s)$$

angegeben, also werden wir die Regression $\mathbb{E}(Y|X)$ benutzen.

Was bedeutet die *beste* Approximation?

Die Variable "weißt alles" über Y, was möglich ist, wenn etwas über X bekannt ist:

Bedingter Erwartungswert

Behauptung. Seien X und Y Zufallsvariablen, wo X diskret und $\mathbb{E}(Y)$ endlich ist. Nehmen wir an, dass $\mathbb{P}(X \leq s)$ positiv ist, dann gilt

$$\mathbb{E}(g(X) \mid X \leq s) = \mathbb{E}(Y \mid X \leq s),$$

wo $g(X) = \mathbb{E}(Y|X)$.

Beweis. Die Transformationsformel für den Erwartungswert gibt

$$\begin{aligned}
\mathbb{E}(g(X) \mid X \leq s) &= \sum_{k \in \operatorname{ran} X} g(k) \cdot \mathbb{P}(X = k \mid X \leq s) \\
&= \sum_{k \leq s} g(k) \cdot \frac{\mathbb{P}(X = k)}{\mathbb{P}(X \leq s)} \\
&= \sum_{k \leq s} \mathbb{E}(Y \mid X = k) \cdot \frac{\mathbb{P}(X = k)}{\mathbb{P}(X \leq s)}
\end{aligned}$$

Bedingter Erwartungswert

Beweis.

$$\begin{aligned}
\mathbb{E}(g(X) \mid X \leq s) &= \sum_{k \leq s} \mathbb{E}(Y \mid X = k) \cdot \frac{\mathbb{P}(X = k)}{\mathbb{P}(X \leq s)} \\
&= \sum_{k \leq s} \frac{1}{\mathbb{P}(X = k)} \mathbb{E}(Y \mathbb{1}_{\{X=k\}}) \cdot \frac{\mathbb{P}(X = k)}{\mathbb{P}(X \leq s)} \\
&= \frac{1}{\mathbb{P}(X \leq s)} \mathbb{E}\left(Y \sum_{k \leq s} \mathbb{1}_{\{X=k\}} \right) \\
&= \frac{1}{\mathbb{P}(X \leq s)} \mathbb{E}(Y \mathbb{1}_{\{X \leq s\}}) = \mathbb{E}(Y \mid X \leq s).
\end{aligned}$$

Bedingter Erwartungswert (stetiger Fall)

Wie können wir diese Approximation konstruieren, falls die Variable X stetig ist?

Problem: wenn X stetig ist, dann gilt $\mathbb{P}(X = s) = 0$ für alle $s \in \mathbb{R}$, also sind die begindten Wahrscheinlichkeiten $\mathbb{P}(A \mid X = s)$ und folglich der bedingte Erwartungswert $\mathbb{E}(Y \mid X = s)$ nicht sinvoll.

Erinnerung: Für eine stetige Variable X können wir statt $\mathbb{P}(X = t)$ den Quotient

$$\frac{\mathbb{P}(t - \varepsilon < X < t + \varepsilon)}{2\varepsilon}$$

betrachten und dann ε gegen 0 schicken. Der Grenzwert (wenn exisitiert) gibt ein Gewicht $f_X(t)$ für die Zahl t und dann bestimmt die Dichtefunktion f_X die Verteilung von X.

Bedingter Erwartungswert (stetiger Fall)

Sei (X, Y) ein stetiger Zufallsvektor mit Verteilungsfunktion $F_{X,Y}$ und Dichtefunktion $f_{X,Y}$. Anstatt $\mathbb{P}(Y \leq t \mid X = s)$ betrachten wir zuerst die Wahrscheinlichkeit

$$\begin{aligned}\mathbb{P}(Y \leq t \mid s-\varepsilon < X < s+\varepsilon) &= \frac{\mathbb{P}(s-\varepsilon < X < s+\varepsilon, Y \leq t)}{\mathbb{P}(s-\varepsilon < X < s+\varepsilon)} \\ &= \frac{F_{X,Y}(s+\varepsilon, t) - F_{X,Y}(s-\varepsilon, t)}{F_X(s+\varepsilon) - F_X(s-\varepsilon)},\end{aligned}$$

wo $F_X(s)$ der Verteilungsfunktion von X ist. Hier nehmen wir natürlich an, dass $\mathbb{P}(s-\varepsilon < X < s+\varepsilon) > 0$ gilt, und falls F_X in s differenzierbar ist, $F_X'(s) \neq 0$, und die partielle Ableitung $\frac{\partial F_{X,Y}}{\partial x}(x, y)$ im Punkt (s, t) existiert, dann gilt

$$\lim_{\varepsilon \to 0} \mathbb{P}(Y \leq t \mid s-\varepsilon < X < s+\varepsilon) = \frac{\frac{\partial F_{X,Y}}{\partial x}(s, t)}{F_X'(s)}.$$

Bedingter Erwartungswert (stetiger Fall)

$$\lim_{\varepsilon \to 0} \mathbb{P}(Y \leq t \mid s - \varepsilon < X < s + \varepsilon) = \frac{\frac{\partial F_{X,Y}}{\partial x}(s,t)}{F_X'(s)}.$$

- Wir werden diesen Grenzwert als die Definizion der bedingten Warhscheinlichkeit $\mathbb{P}(Y \leq t \mid X = s)$ benutzen.
- Im Nenner auf der rechten Seite erscheint der Wert der Dichtefunktion $f_X(s) = F_X'(s)$ von X an der Stelle s.
- Man kann auch überlegen, dass

$$\frac{\partial F_{X,Y}}{\partial x}(s,t) = \int_{-\infty}^{t} f_{X,Y}(s,y)\, dy$$

gilt, wenn zum Beispiel $f_{X,Y}$ stetig ist. Also geben wir die folgende Definition:

$$\mathbb{P}(Y \leq t \mid X = s) := \int_{-\infty}^{t} \frac{f_{X,Y}(s,y)}{f_X(s)}\, dy.$$

Bedingter Erwartungswert (stetiger Fall)

Bemerken wir, dass

$$f_X(s) = \int_{-\infty}^{\infty} f_{X,Y}(s,y)\,dy$$

gilt, und deshalb ist die Funktion

$$\frac{f_{X,Y}(s,\cdot)}{f_X(s)}$$

für alle $s \in \mathbb{R}$ mit $f_X(s) > 0$ eine Dichtefunktion in einer Variablen, also gibt sie wirklich ein Wahrscheinlichkeitsmaß.

Bedingter Erwartungswert (stetiger Fall)

Definition. Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{X,Y}$. Dann ist *die bedingte Dichtefunktion von Y auf X*

$$f_{Y|X}(t \mid s) = \frac{f_{X,Y}(s,t)}{f_X(s)} = \frac{f_{X,Y}(s,t)}{\int_{-\infty}^{\infty} f_{X,Y}(s,y)\,dy}$$

falls $f_X(s) \neq 0$ ist, und $f_{Y|X}(t \mid s) = 0$ sonst.

Wenn $f_X(s) \neq 0$, dann ist $f_{Y|X}(t \mid s)$ eine Dichtefunktion, und wir definieren den bedingten Erwartungswert

$$\mathbb{E}(Y \mid X = s) := \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid s)\,dy$$

als der zugehörige Erwartungswert. Dieser Erwartungswert ist endlich, falls $\mathbb{E}(Y)$ endlich ist, und die obige Definition ist sinnvoll auch wenn $f_X = 0$, dann wird $\mathbb{E}(Y \mid X = s)$ einfach Null.

Bedingter Erwartungswert (stetiger Fall)

Beispiel. Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion

$$f_{X,Y}(s,t) = \begin{cases} 15s^2 t, & \text{falls } 0 < s < t < 1, \\ 0 & \text{sonst.} \end{cases}$$

Wir bestimmen den bedingten Erwartungswert $\mathbb{E}(Y \mid X = s)$. Die Dichtefunktion von X ist

$$f_X(s) = \int_{-\infty}^{\infty} f_{X,Y}(s,y)\,dy = \int_s^1 15s^2 y\,dy = \frac{15}{2}(s^2 - s^4),$$

falls $0 < s < 1$ gilt, und 0 ansonsten. Dann ist

$$f_{Y|X}(t \mid s) = \frac{f_{X,Y}(s,t)}{f_X(s)} = \frac{15s^2 t}{\frac{15}{2}(s^2 - s^4)} = \frac{2t}{1 - s^2}$$

falls $0 < s < t < 1$, und $f_{Y|X}(t \mid s) = 0$ sonst.

Bedingter Erwartungswert (stetiger Fall)

Beispiel. Also

$$\mathbb{E}(Y \mid X = s) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid s)\, dy = \int_s^1 y \cdot \frac{2y}{1 - s^2}\, dy$$

$$= \frac{1}{1 - s^2} \int_s^1 2y^2\, dy = \frac{2}{3} \cdot \frac{1 - s^3}{1 - s^2} = \frac{2}{3} \cdot \frac{1 + s + s^2}{1 + s},$$

falls $0 < s < 1$, und Null sonst.

Bedingter Erwartungswert (stetiger Fall)

Definition. Sei (X, Y) ein stetiger Zufallsvektor, wo $\mathbb{E}(Y)$ endlich ist. Dann ist die Funktion

$$g(s) := \mathbb{E}(Y \mid X = s)$$

(als eine Funktion von s) die sogenannte *Regressionsfunktion von Y auf X*, und die Variable $g(X)$ ist die (stetige) *Regression von Y auf X*.
Notation: $g(X) = \mathbb{E}(Y \mid X)$.

Im obigen Beispiel war

$$\mathbb{E}(Y \mid X = s) = \frac{2}{3} \cdot \frac{1 + s + s^2}{1 + s},$$

deshalb ist die Regression von Y auf X

$$\frac{2}{3} \cdot \frac{1 + X + X^2}{1 + X}.$$

Bedingter Erwartungswert (stetiger Fall)

Wie im diskreten Fall, die Regression wird eine gute Approximation von Y anhand der Variablen X:

Behauptung. Sei (X, Y) ein stetiger Zufallsvektor, wo $\mathbb{E}(Y)$ endlich ist. Nehmen wir an, dass $\mathbb{P}(X \leq s)$ positiv ist, dann gilt

$$\mathbb{E}(g(X) \mid X \leq s) = \mathbb{E}(Y \mid X \leq s),$$

wo $g(X) = \mathbb{E}(Y|X)$. Hier ist also $g : \mathbb{R} \to \mathbb{R}$ die Regressionsfunktion

$$s \mapsto \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid s)\, dy.$$

Bedingter Erwartungswert (stetiger Fall)

Beweis. Es gilt

$$\mathbb{E}(g(X) \mid X \leq s) = \frac{1}{\mathbb{P}(X \leq s)} \mathbb{E}(g(X) \mathbb{1}_{\{X \leq s\}}),$$

und

$$\mathbb{E}(g(X) \mathbb{1}_{\{X \leq s\}}) = \int_{-\infty}^{\infty} g(x) \mathbb{1}_{\{X \leq s\}}(x) \cdot f_X(x)\, dx = \int_{-\infty}^{s} g(x) f_X(x)\, dx$$

Falls $f_X(x) > 0$, dann gilt

$$\begin{aligned} g(x) f_X(x) &= \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\, dy \cdot f_X(x) \\ &= \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(x,y)}{f_X(x)}\, dy \cdot f_X(x) \\ &= \int_{-\infty}^{\infty} y \cdot f_{X,Y}(x,y)\, dy. \end{aligned}$$

Bedingter Erwartungswert (stetiger Fall)

Beweis. Falls $f_X(x) = 0$, dann ist auch $g(x)f_X(x) = 0$. Aber dann gilt

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = 0,$$

und es folgt, dass

$$\int_{-\infty}^{\infty} y \cdot f_{X,Y}(x,y)\,dy = 0$$

(wir beweisen das nicht). Also gilt immmer

$$g(x)f_X(x) = \int_{-\infty}^{\infty} y \cdot f_{X,Y}(x,y)\,dy.$$

Bedingter Erwartungswert (stetiger Fall)

Beweis. Das heißt,

$$\begin{aligned}
\mathbb{E}(g(X) \mid X \leq s) &= \frac{1}{\mathbb{P}(X \leq s)} \int_{-\infty}^{s} \int_{-\infty}^{\infty} y \cdot f_{X,Y}(x,y)\, dy\, dx \\
&= \frac{1}{\mathbb{P}(X \leq s)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_{\{X \leq s\}}(x) \cdot y \cdot f_{X,Y}(x,y)\, dy\, dx \\
&= \frac{1}{\mathbb{P}(X \leq s)} \mathbb{E}(\mathbb{1}_{\{X \leq s\}} Y) = \mathbb{E}(Y \mid X \leq s).
\end{aligned}$$

Bedingter Erwartungswert

Die Regression kann auch im Allgemeinfall definiert werden, indem die obige Behauptung als Definition benutzt wird:

Definition. Seien X und Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Die Regression von Y auf X ist die Zufallsvariable $g(X)$, für die die Gleichuing

$$\mathbb{E}(g(X) \mid X \leq s) = \mathbb{E}(Y \mid X \leq s)$$

für alle $s \in \mathbb{R}$ mit $\mathbb{P}(X \leq s) > 0$ gilt.

Notation: $g(X) = \mathbb{E}(Y \mid X)$.

Bedingter Erwartungswert

Bemerkungen.

- Die Zufallsvariable $\mathbb{E}(Y \mid X)$, wenn sie existiert, ist nicht eindeutig bestimmt. Mann kann zum Beispiel den Wert von g an endlich vielen Stellen verändern, wenn X diese Werte nicht mit einer positiven Wahrscheinlichkeit annimmt.
- Die Funktion g heißt die *Regressionsfunktion*. Notation:

$$s \mapsto \mathbb{E}(Y \mid X = s).$$

  Diese Notation ist nicht präzis: der bedingte Erwartungswert $\mathbb{E}(Y \mid X = s)$ wurde im Allgemeinen nicht definiert, und g kann zum Beispiel auf der Menge $\mathbb{R} \setminus \operatorname{ran} X$ beliebige Werte annehmen, ohne die Verteilung von $g(X)$ zu beeinflussen.
- Falls $\mathbb{E}(Y)$ endlich ist, dann existiert $\mathbb{E}(Y \mid X)$.

Bedingter Erwartungswert

Bemerkung. Offensichtlich gelten

- $\mathbb{E}(X \mid X) = X$ falls $\mathbb{E}(X)$ endlich ist,
- $\mathbb{E}(c \mid X) = c$ für alle $c \in \mathbb{R}$.

Bedingter Erwartungswert

Behauptung. Seien X, Y und Z Zufallsvariablen, wo $\mathbb{E}(Y)$ und $\mathbb{E}(Z)$ endlich sind. Dann gelten

(i) $\mathbb{E}(aY + bZ \mid X) = a\mathbb{E}(Y \mid X) + b\mathbb{E}(Z|X)$ für alle $a, b \in \mathbb{R}$,

(ii) $\mathbb{E}(h(X)Y \mid X) = h(X)\mathbb{E}(Y \mid X)$ für jede stetige Funktion h,

(iii) $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$, falls X und Y unabhängig sind.

Bedingter Erwartungswert

Die Regression ist auch die "beste" Approximation:

Behauptung. Seien X, Y Zufallsvariablen, wo $\mathbb{D}(Y)$ endlich ist. Dann ist der Erwartungswert

$$\mathbb{E}((Y - g(X))^2)$$

genau dann minimal, falls $\mathbb{P}(g(X) = \mathbb{E}(Y \mid X)) = 1$ gilt.

Bedingter Erwartungswert

Satz. (Satz vom totalen Erwartungswert)
Seien X, Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Dann gilt

$$\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y).$$

Diskreter Fall:

Falls X diskret ist, und

$$S_X = \{s \in \mathbb{R} : \mathbb{P}(X = s) > 0\},$$

dann gilt

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_{s \in S_X} \mathbb{E}(Y \mid X = s)\mathbb{P}(X = s),$$

wo $g(s) = \mathbb{E}(Y \mid X = s)$ die Regressionsfunktion von Y auf X ist.

Bedingter Erwartungswert

Satz. (Satz vom totalen Erwartungswert)
Seien X, Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Dann gilt

$$\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y).$$

Im stetigen Fall:

Seien X, Y Zufallsvariablen, wo $\mathbb{E}(Y)$ endlich ist. Falls X stetig mit Dichtefunktion f_X ist, dann gilt

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = s) f_X(s)\, ds,$$

wo $\mathbb{E}(Y \mid X = s)$ die Regressionsfunktion von Y auf X ist.

Bedingte Wahrscheinlichkeit im stetigen Fall

Erinnerung. Der Satz der totalen Wahrscheinlichkeit wurde nur für vollständige Ereignissysteme von Ereignissen positiver Wahrscheinlichkeit betrachtet:

Satz. (Satz der totalen Wahrscheinlichkeit)
Sei $A_1, \ldots, A_n \in \mathcal{F}$ ein vollständiges Ereignissystem mit $\mathbb{P}(A_i) > 0$. Dann gilt

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

für alle Ereignisse $B \in \mathcal{F}$.

Bedingte Wahrscheinlichkeit im stetigen Fall

Erinnerung. Ähnlich kann der Satz vom totalen Erwartungswert für vollständige Ereignissysteme formuliert werden:

Satz. (Satz vom totalen Erwartungswert)
Sei Y eine Zufallsvariable mit endlichem Erwartungswert. Sei noch $A_1, \dots, A_n$ ein vollständiges Ereignissystem mit $\mathbb{P}(A_i) > 0$ für alle $1 \leq i \leq n$. Dann gilt

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y \mid A_i)\mathbb{P}(A_i).$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Bemerkung. Der erste Satz folgt aus dem zweiten:

Sei B ein Ereignis, und sei $\mathbb{1}_B$ die Indikatorvariable von B. Sei noch $A_1, \ldots, A_n$ ein vollständiges Ereignissystem mit $\mathbb{P}(A_i) > 0$ für alle $1 \leq i \leq n$. Dann gilt wegen des Satzes vom totalen Erwartungswert

$$\mathbb{P}(B) = \mathbb{E}(\mathbb{1}_B) = \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_B \mid A_i)\mathbb{P}(A_i).$$

Nach seiner Definition ist der Wert des bedingten Erwartungswertes $\mathbb{E}(\mathbb{1}_B \mid A_i)$

$$0 \cdot \mathbb{P}(\mathbb{1}_B = 0 \mid A_i) + 1 \cdot \mathbb{P}(\mathbb{1}_B = 1 \mid A_i) = \mathbb{P}(B \mid A_i),$$

also folgt der Satz der totalen Wahrscheinlichkeit.

Bedingte Wahrscheinlichkeit im stetigen Fall

Bemerkung. Der Satz vom totalen Erwartungswert kann auch für eine diskrete Zufallsvariable formuliert werden, und daraus folgt der Satz der totalen Wahrscheinlichkeit wie oben:

Sei B ein Ereignis, und sei $\mathbb{1}_B$ die Indikatorvariable von B. Sei noch X eine diskrete Zufallsvariable, und

$$S_X = \{s \in \mathbb{R} : \mathbb{P}(X = s) > 0\}.$$

Dann gilt wegen des Satzes vom totalen Erwartungswert

$$\begin{aligned} \mathbb{P}(B) = \mathbb{E}(\mathbb{1}_B) &= \sum_{s \in S_X} \mathbb{E}(\mathbb{1}_B \mid X = s)\mathbb{P}(X = s) \\ &= \sum_{s \in S_X} \mathbb{P}(B \mid X = s)\mathbb{P}(X = s). \end{aligned}$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Bemerkung. Zwar die bedingte Wahrscheinlichkeit $\mathbb{P}(Y \leq t \mid X = s)$ wurde für stetige Zufallsvektoren (X, Y) definiert, im Allgemeinen ist der Ausdruck $\mathbb{P}(B \mid X = s)$ nicht sinnvoll, falls $\mathbb{P}(X = s) = 0$ gilt (also zum Beispiel falls X stetig ist).

Aber die Regressionsfunktion $g(s) = \mathbb{E}(\mathbb{1}_B \mid X = s)$ wurde auch im Allgemeinfall durch die Gleichung

$$\mathbb{E}(g(X) \mid X \leq s) = \mathbb{E}(\mathbb{1}_B \mid X \leq s)$$

definiert, und kann als Definizion für $\mathbb{P}(B \mid X = s)$ dienen.

Bedingte Wahrscheinlichkeit im stetigen Fall

Definition. Sei B ein Ereignis, und sei X eine stetige Zufallsvariable. Dann ist *die bedingte Wahrscheinlichkeit von* B *auf die Variable* X (a B-nek az X-re vett feltételes valószínűsége) ist durch

$$s \mapsto \mathbb{E}(\mathbb{1}_B \mid X = s)$$

definiert. Notation: $\mathbb{P}(B \mid X = s)$.

$\mathbb{P}(B \mid X = s)$ gibt eine Approximation für die Wahrscheinlichkeit von B, wenn der Wert von X "in der Nähe" von s liegt.

Bedingte Wahrscheinlichkeit im stetigen Fall

Bemerkung. Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion $f_{X,Y}(s,t)$. Dann ist die bedingte Dichtefuntion von Y auf X

$$f_{Y|X}(t \mid s) = \begin{cases} \frac{f_{X,Y}(s,t)}{f_X(s)}, & \text{falls } f_X(s) \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

In diesem Fall stimmen die neue und die alte Definition von $\mathbb{P}(Y \leq t \mid X = s)$ überein, also gilt

$$\mathbb{P}(Y \leq t \mid X = s) = \int_{-\infty}^{t} f_{Y|X}(y \mid s)\, dy.$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Falls die Variable X stetig ist, dann gibt der Satz vom totalen Erwartungswert:

Satz. (Satz der totalen Wahrscheinlichkeit - stetiger Fall)
Sei B ein Ereignis, und sei X eine stetige Zufallsvariable mit Dichtefunktion f_X. Dann gilt

$$\mathbb{P}(B) = \int_{-\infty}^{\infty} \mathbb{P}(B \mid X = y) f_X(y)\, dy.$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Beispiel. Sei X die Zeitdauer (in Stunden), die ein:e Student:in mit der Vorbereitung auf eine Prüfung verbringt. Nehmen wir an, dass X uniformverteilt auf dem Intervall $[\varepsilon, 20]$ ist, wo ε eine positive Zahl ist.

Wenn der:die Student:in s Stunden mit der Vorbereitung verbringt, dann wird die Note mit Wahrscheinlichkeit $\left(\frac{s}{21}\right)^2$ eine 5. Was ist die Wahrscheinlichkeit dafür, dass der:die Student:in eine 5 bekommt?

Bedingte Wahrscheinlichkeit im stetigen Fall

Die Dichtefunktion von X ist

$$f_X(s) = \begin{cases} \frac{1}{20-\varepsilon}, & \text{falls } \varepsilon \leq s \leq 20, \\ 0 & \text{sonst.} \end{cases}$$

Wir haben auch

$$\mathbb{P}(\text{die Note ist } 5 \mid X = s) = \left(\frac{s}{21}\right)^2,$$

also

$$\mathbb{P}(\text{die Note ist } 5) = \int_{-\infty}^{\infty} \mathbb{P}(\text{die Note ist } 5 \mid X = y) f_X(y)\, dy$$

$$= \int_{\varepsilon}^{20} \left(\frac{y}{21}\right)^2 \frac{1}{20-\varepsilon}\, dy$$

Bedingte Wahrscheinlichkeit im stetigen Fall

$$\begin{aligned}\mathbb{P}(\text{die Note ist } 5) &= \int_{\varepsilon}^{20} \left(\frac{y}{21}\right)^2 \frac{1}{20-\varepsilon}\, dy \\ &= \left[\frac{y^3}{3 \cdot 21^2 \cdot (20-\varepsilon)}\right]_{\varepsilon}^{20} \\ &= \frac{\varepsilon^2 + 20\varepsilon + 20^2}{3 \cdot 21^2}.\end{aligned}$$

Falls $\varepsilon = 1$ gilt, dann ist diese Wahrscheinlichkeit ungefähr 0,3182.

Bedingte Wahrscheinlichkeit im stetigen Fall

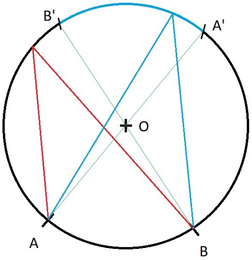
Beispiel. Wir wählen 3 Punkte A, B und C zufällig und unabhänging voneinander auf der Einheitskreislinie. Was ist die Wahrscheinlichkeit dafür, dass der Mittelpunkt O des Kreises im Dreieck $ABC\triangle$ enthalten ist?

Idee: wenn die Punkte A und B schon bekannt sind, dann ist es leicht, diese Wahrscheinlichkeit zu bestimmen.

Bedingte Wahrscheinlichkeit im stetigen Fall

Seien nämlich A' bzw. B' die Bilder von A bzw. B bei der Spiegelung am Punkt O. Der Mittelpunkt wird denau dann im Dreieck $ABC\triangle$, wenn C auf dem kurzeren Kreisbogen zwischen A' und B' liegt.

Sei $\lambda(AB)$ die Länge des kurzeren Kreisbogens zwischen A und B, dann ist die gefragte Wahrscheinlichkeit $\frac{\lambda(AB)}{2\pi}$ (wenn A und B schon bekannt sind).



Bedingte Wahrscheinlichkeit im stetigen Fall

Die Funktion $\lambda(AB)$ ist eine stetige Zufallsvariable mit Bildmenge $[0;\pi]$. Sei f_λ ihre Dichtefunktion, und sei noch E das Ereignis, dass der Mittelpunkt im $ABC\triangle$ enthalten ist. Dann gilt wegen des Satzes der totalen Wahrscheinlichkeit

$$\mathbb{P}(E) = \int_{-\infty}^{\infty} \mathbb{P}(E \mid \lambda(AB) = y) f_\lambda(y)\, dy$$

$$= \int_0^{\pi} \frac{y}{2\pi} f_\lambda(y)\, dy = \frac{\mathbb{E}(\lambda(AB))}{2\pi}.$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Es bleibt, $\mathbb{E}(\lambda(AB))$ zu bestimmen.

- $\lambda(AB)$ ist die Bogenlänge zwischen zwei uniform zufällig und unabhängig gewählten Punkten.
- Aber auch A' und B sind ein Paar von uniform zufällig und unabhängig gewählten Punkten.
- Also sind $\lambda(AB)$ und $\lambda(A'B)$ identisch verteilt.
- Es gilt noch $\lambda(AB) + \lambda(A'B) = \pi$, also

$$\begin{aligned}\pi = \mathbb{E}(\pi) &= \mathbb{E}(\lambda(AB) + \lambda(A'B)) \\ &= \mathbb{E}(\lambda(AB)) + \mathbb{E}(\lambda(A'B)) = 2\mathbb{E}(\lambda(AB))\end{aligned}$$

- Das heisst:

$$\mathbb{E}(\lambda(AB)) = \frac{\pi}{2}, \qquad \mathbb{P}(E) = \frac{\pi/2}{2\pi} = \frac{1}{4}.$$

Bedingte Wahrscheinlichkeit im stetigen Fall

Aufgabe. Immer wenn Onkel Otto einen Stab zerbricht, die Bruchstelle wird ein uniform verteilter Punkt auf dem mittleren Drittel des Stabs.
Nehmen wir an, dass Onkel Otto einen 15 Zentimeter langen Stab zerbricht, und danach zebricht er auch das Stück, das in seiner linken Hand bleibt. Geben wir die Dichtefunktion der Länge des Stücks an, das am Ende (nach dem zweiten Bruch) in seiner linken Hand bleibt.

Mehrdimensionale Normalverteilung

Wir verallgemeinern die Normalverteilung folgenderweise:

- Der Fehler einer Messung kann durch eine normalverteilte Zufallsvariable beschreibt werden.
- Nehmen wir an, dass das Ergebnis der Messung eine Position mit 2 Koordinaten ist. Zum Beispiel: wir möchten die Position einer Senderanlage bestimmen.
- Was für Eigenschaften sollte die Verteilung des Fehlers der Messung - also die Verteilung des Differenzvektors des gemissten Ergebnispunktes und der tatsächlichen Position - im idealen Fall haben?

Mehrdimensionale Normalverteilung

Seien X und Y die Koordinaten dieser Differenz. Wir nehmen die folgenden Eigenschaften an:

- die Verteilung des Zufallsvektors $(X, Y)^T$ ist stetig,
- die Dichtefunktion $f_{X,Y}$ ist drehsymmetrisch, also hängt der Wert $f_{X,Y}(s,t)$ nur von der Länge des Vektors $(s,t)^T$ ab, d.h.:

  $$f_{X,Y}(s,t) = h(s^2 + t^2)$$

  gilt für eine Funktion $h : [0, \infty) \to \mathbb{R}$,
- und die Koordinaten X und Y sind unabhängig, also gilt

  $$f_{X,Y}(s,t) = f_X(s) f_Y(t).$$

Mehrdimensionale Normalverteilung

Bemerkung. Bei diesem Thema werden wir die Elemente von $\mathbb{R}^n$ als Spaltenvektoren betrachten. Wenn wir mit Zeilenvektoren arbeiten möchten, dann werden wir die transponierten Vektoren nützen.

Mehrdimensionale Normalverteilung

Behauptung. Sei $(X, Y)^T$ ein Zufallsvektor mit den obigen Eigenschaften, dann gibt es reelle Zahlen $a, c \in \mathbb{R}$, wo a negativ ist, und $f_{X,Y}(s,t) = e^{a(s^2+t^2)-c}$ gilt.

Die Parameter a und c können so gewählt werden, dass wir die sogenannte Standardnormalverteilung bekommen. Im Allgemeinen, im n-dimensionalen Fall:

Defintion. Der Zufallsvektor $\underline{X} = (X_1, \ldots, X_n)^T$ ist *standardnormalverteilt*, falls seine Dichtefunktion durch

$$f(t_1, \ldots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2}$$

angegeben ist.

Mehrdimensionale Normalverteilung

Defintion. Der Zufallsvektor $\underline{X} = (X_1, \ldots, X_n)^T$ ist *standardnormalverteilt*, falls seine Dichtefunktion durch

$$f(t_1, \ldots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2}$$

gegeben ist.

Es gilt offenbar

$$f(t_1, \ldots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{t_i^2}{2}} = \prod_{i=1}^{n} f_{X_i}(t_i),$$

wo f_{X_i} die Dichtefunktion von X_i ist. Also sind die Koordinaten eines standardnormalverteilten Zufallsvektors unabhängig.

Mehrdimensionale Normalverteilung

Defintion. Der Zufallsvektor $\underline{Y} = (Y_1, \ldots, Y_n)^T$ ist *normalverteilt*, falls es einen standardnormalverteilten Zufallsvektor $\underline{X} = (X_1, \ldots, X_n)^T$, eine Matrix $A \in \mathbb{R}^{n \times n}$ und einen Vektor $\underline{\mu} \in \mathbb{R}^n$ gibt, für die

$$\underline{Y} = A\underline{X} + \underline{\mu}$$

gilt. Die Verteilung von $\underline{Y}$ heißt *singulär* (elfajuló), falls A singulär ist (d.h. $\det A = 0$).

Um diese Verteilungen zu studieren, benötigen wir einige Begriffe:

Mehrdimensionale Normalverteilung

Definition. Sei $\underline{Y} = (Y_1, \dots, Y_n)^T$ ein (beliebiger) Zufallsvektor, dann ist der Vektor $(\mathbb{E}(Y_1), \dots, \mathbb{E}(Y_n))^T$ (falls er existiert) der *Erwartungswertvektor* von $\underline{Y}$ ($\underline{Y}$ várhatóérték-vektora).

Bezeichnung: $\mathbb{E}\underline{Y}$

Definition. Sei $\underline{Y} = (Y_1, \dots, Y_n)^T$ ein (beliebiger) Zufallsvektor, dann heißt die Matrix

$$\mathrm{cov}(\underline{Y}) = \begin{pmatrix} \mathrm{cov}(Y_1, Y_1) & \mathrm{cov}(Y_1, Y_2) & \dots & \mathrm{cov}(Y_1, Y_n) \\ \mathrm{cov}(Y_2, Y_1) & \mathrm{cov}(Y_2, Y_2) & \dots & \mathrm{cov}(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(Y_n, Y_1) & \mathrm{cov}(Y_n, Y_2) & \dots & \mathrm{cov}(Y_n, Y_n) \end{pmatrix}$$

(falls sie existiert) die *Kovarianzmatrix* von $\underline{Y}$.

Mehrdimensionale Normalverteilung

Bemerkung. Die Kovarianzmatrix eines Zufallsvektors ist symmetrisch, weil die Kovarianz symmetrisch ist, also gilt $\mathrm{cov}\,(Y_i, Y_j) = \mathrm{cov}\,(Y_j, Y_i)$ für alle $1 \leq i, j \leq n$.

Mehrdimensionale Normalverteilung

Bemerkung. Die Kovarianzmatrix ist positiv semidefinit, also gilt

$$\underline{x}^T \cdot \operatorname{cov}(\underline{Y}) \cdot \underline{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \cdot \operatorname{cov}(Y_i, Y_j) \cdot x_j \geq 0$$

für alle $\underline{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Nämlich gilt wegen der Bilinearität der Kovarianz

$$\begin{aligned} 0 \leq \mathbb{D}^2\left(\sum_{i=1}^{n} x_i Y_i\right) &= \operatorname{cov}\left(\sum_{i=1}^{n} x_i Y_i, \sum_{j=1}^{n} x_j Y_j\right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \cdot \operatorname{cov}(Y_i, Y_j) \cdot x_j . \end{aligned}$$

Wir erhalten auch, dass diese Summe genau dann Null ist, wenn $\sum_{i=1}^{n} x_i Y_i$ mit Wahrscheinlichkeit 1 konstant ist.

Mehrdimensionale Normalverteilung

Ergänzung. Die Kovarianzmatrix ist genau dann positiv definit, d.h.:

$$\underline{x}^T \cdot \operatorname{cov}(\underline{Y}) \cdot \underline{x} > 0$$

gilt genau dann für alle $0 \neq \underline{x} \in \mathbb{R}^n$, wenn $\operatorname{cov}(\underline{Y})$ nicht singulär ist, also wenn $\det[\operatorname{cov}(\underline{Y})] \neq 0$ gilt.

Mehrdimensionale Normalverteilung

Behauptung. Sei $\underline{Y} \in \mathbb{R}^n$ ein Zufallsvektor. Dann gilt

$$\mathrm{cov}\,(\underline{Y}) = \mathbb{E}\left[(\underline{Y} - \mathbb{E}(\underline{Y}))(\underline{Y} - \mathbb{E}(\underline{Y}))^T\right],$$

falls die Kovarianzmatrix existiert.

Hier ist $(\underline{Y} - \mathbb{E}(\underline{Y}))$ ein Spaltenvektor und $(\underline{Y} - \mathbb{E}(\underline{Y}))^T$ ist ein Zeilenvektor, so ist ihr Produkt eine $n \times n$-Matrix von Zufallsvariablen. Der außere Erwartungswert bedeutet, dass die Matrix der elementweisen Erwartungswerte genommen wird.

Die Behauptung folgt dann sofort aus der Definizion der Kovarianz von 2 Variablen.

Mehrdimensionale Normalverteilung

Behauptung. Sei $\underline{X} = (X_1, \ldots, X_n)^T$ ein standardnormalverteilter Zufallsvektor, und $\underline{Y} = A\underline{X} + \underline{\mu}$, wo $A \in \mathbb{R}^{n\times n}$ und $\underline{\mu} \in \mathbb{R}^n$. Dann gelten $\underline{\mu} = \mathbb{E}\underline{Y}$ und $\mathrm{cov}(\underline{Y}) = AA^T$.

Beweis: Seien $A = (a_{ij})_{1\le i,j\le n}$ und $\underline{\mu} = (\mu_1, \ldots, \mu_n)^T$. Dann ist die $i$te Koordinate von $\underline{Y}$

$$a_{i1}X_1 + \cdots + a_{in}X_n + \mu_i.$$

Hier sind die Variablen X_i standardnormalverteilt, also gilt $\mathbb{E}(X_i) = 0$ für alle $1 \le i \le n$. Dann ergibt sich die Gleichung

$$\mathbb{E}(a_{i1}X_1 + \cdots + a_{in}X_n + \mu_i) = \sum_{j=1}^{n} a_{ij}\mathbb{E}(X_j) + \mu_i = \mu_i,$$

und die erste Behauptung folgt.

Mehrdimensionale Normalverteilung

Beweis: Die Kovarianzmatrix ist durch

$$\begin{aligned}\operatorname{cov}(\underline{Y}) &= \mathbb{E}\left[(\underline{Y}-\underline{\mu})(\underline{Y}-\underline{\mu})^T\right] \\ &= \mathbb{E}\left[(A\underline{X})(\underline{X}^T A^T)\right] = \mathbb{E}\left[A(\underline{XX}^T)A^T\right]\end{aligned}$$

gegeben. Hier ist

$$\underline{XX}^T = \begin{pmatrix} X_1^2 & X_1X_2 & \dots & X_1X_n \\ X_2X_1 & X_2^2 & \dots & X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ X_nX_1 & X_nX_2 & \dots & X_n^2 \end{pmatrix},$$

also sind die Elemente der Kovarianzmatrix Linearkombinationen der Variablen X_iX_j $(1 \leq i, j, \leq n)$.

Mehrdimensionale Normalverteilung

Beweis: Wegen der Linearität des Erwartungswertes erhalten wir dann

$$\operatorname{cov}(\underline{Y}) = A \begin{pmatrix} \mathbb{E}(X_1^2) & \mathbb{E}(X_1 X_2) & \dots & \mathbb{E}(X_1 X_n) \\ \mathbb{E}(X_2 X_1) & \mathbb{E}(X_2^2) & \dots & \mathbb{E}(X_2 X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(X_n X_1) & \mathbb{E}(X_n X_2) & \dots & \mathbb{E}(X_n^2) \end{pmatrix} A^T.$$

Wegen der Unabhängigkeit von $X_1, \dots, X_n$ gilt

$$\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$$

falls $i \neq j$, und endlich haben wir auch

$$\mathbb{E}(X_i^2) = \mathbb{D}^2(X_i) + \mathbb{E}(X_i)^2 = 1$$

für alle $1 \leq i \leq n$.

Mehrdimensionale Normalverteilung

Beweis: Dann folgt

$$\operatorname{cov}(\underline{Y}) = A \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} A^T = AA^T.$$

Mehrdimensionale Normalverteilung

Anhand dieser Objekte kann auch die Dichtefunktion von $\underline{Y}$ ausgedrückt werden:

Behauptung. Sei $\underline{Y} = (Y_1, \ldots, Y_n)^T$ ein nicht singulärer n-dimensionaler normalverteilter Zufallsvektor mit dem Erwartungswertvektor $\underline{\mu} = \mathbb{E}\underline{Y}$ und mit der Kovarianzmatrix Σ. Dann gilt $\det\Sigma \neq 0$, und die Dichtefunktion von $\underline{Y}$ ist

$$f_{\underline{Y}}(t_1, \ldots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{t}-\underline{\mu})^T \Sigma^{-1} (\underline{t}-\underline{\mu})}.$$

Bemerkung. Wegen der obigen Behauptung bestimmen die Kovarianzmatrix Σ und der Erwartungswertvektor $\underline{\mu}$ die Verteilung eines nicht singulären normalverteilten Zufallsvektor $\underline{Y}$ eindeutig. Diese Behauptung gilt auch wenn $\underline{Y}$ singulär ist, aber dann hat $\underline{Y}$ keine Dichtefunktion. Man benutzt also die Bezeichnung

$$\underline{Y} \sim N(\underline{\mu}; \Sigma).$$

Mehrdimensionale Normalverteilung

Bemerkung. Die Kovarianzmatrix und der Erwartungswertvektor bestimmen eine nicht singuläre Normalverteilung eindeutig. Aber verschiedene Matrizen A_1, A_2 können dieselbe Kovarianzmatrix geben (falls $A_1 A_1^T = A_2 A_2^T$ gilt).

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Wegen unserer Annahme gilt

$$\underline{Y} = A\underline{X} + \underline{\mu},$$

wo $\underline{\mu} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, $\det A \neq 0$, und der Zufallsvektor $\underline{X}$ standardnormalverteilt ist.

Daraus folgt sofort, dass $\Sigma = \operatorname{cov}(\underline{Y}) = AA^T$, und wegen des Determinantenmultiplikationssatzes gilt deshalb

$$\det \Sigma = \det(AA^T) = (\det A)(\det A^T) = (\det A)^2 \neq 0.$$

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Einem Vektor $\underline{t} = (t_1, \dots, t_n)^T \in \mathbb{R}^n$ ordnen wir die Produktmenge

$$C_{\underline{t}} = (-\infty; t_1) \times \cdots \times (-\infty; t_n)$$

zu. Dann ist der Verteilungsfunktion von $\underline{Y}$ an der Stelle $\underline{t}$

$$F_{\underline{Y}}(\underline{t}) = \mathbb{P}(\underline{Y} \in C_{\underline{t}}) = \mathbb{P}(A\underline{X} + \underline{\mu} \in C_{\underline{t}}) = \mathbb{P}(\underline{X} \in A^{-1}(C_{\underline{t}} - \underline{\mu}))$$

$$= \int_{A^{-1}(C_{\underline{t}} - \underline{\mu})} f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_{\varphi(C_{\underline{t}})} f_{\underline{X}}(\underline{x})\, d\underline{x},$$

wo $f_{\underline{X}}$ die Dichtefunktion von $\underline{X}$ ist, und

$$\varphi : \mathbb{R}^n \to \mathbb{R}^n; \qquad \underline{t} \mapsto A^{-1}(\underline{t} - \underline{\mu})$$

eine stetig differenzierbare und bijektive Funktion ist.

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Dann gilt wegen der mehrdimensionalen Substitutionsregel

$$F_{\underline{Y}}(\underline{t}) = \int_{C_{\underline{t}}} f_{\underline{X}}(\varphi(\underline{x})) \left|\det \varphi'(\underline{x})\right| \, d\underline{x}.$$

Hier bezeichnet $\varphi'(\underline{x})$ die Jacobi-Matrix, die in diesem Fall einfach A^{-1} für alle $\underline{x} \in \mathbb{R}^n$ ist, und deshalb gilt

$$\left|\det \varphi'(\underline{x})\right| = \left|\det A^{-1}\right| .$$

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Sei

$$E_n = E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

die $n \times n$-Einheitsmatrix, dann haben wir

$$1 = \det E = \det(AA^{-1}) = (\det A)(\det A^{-1}),$$

also gilt

$$\left|\det \varphi'(\underline{x})\right| = \left|\det A^{-1}\right| = \frac{1}{|\det A|} = \frac{1}{\sqrt{(\det A)^2}} = \frac{1}{(\det \Sigma)^{\frac{1}{2}}}.$$

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Bisher zeigten wir

$$F_{\underline{Y}}(\underline{t}) = \frac{1}{(\det\Sigma)^{\frac{1}{2}}} \int_{C_{\underline{t}}} f_{\underline{X}}(\varphi(\underline{x}))\, d\underline{x}$$

$$= \frac{1}{(\det\Sigma)^{\frac{1}{2}}} \int_{C_{\underline{t}}} f_{\underline{X}}(A^{-1}(\underline{x}-\underline{\mu}))\, d\underline{x}.$$

Die Dichtefunktion von $\underline{X}$ ist

$$f_{\underline{X}}(\underline{t}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\underline{t}^T\cdot\underline{t})},$$

also folgt

$$F_{\underline{Y}}(\underline{t}) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}} \int_{C_{\underline{t}}} e^{-\frac{1}{2}\left[(A^{-1}(\underline{x}-\underline{\mu}))^T\cdot(A^{-1}(\underline{x}-\underline{\mu}))\right]}\, d\underline{x}.$$

Mehrdimensionale Normalverteilung

Beweis der Behauptung. Endlich gelten

$$\begin{aligned}(A^{-1}(\underline{x}-\underline{\mu}))^T \cdot (A^{-1}(\underline{x}-\underline{\mu})) &= [((\underline{x}-\underline{\mu})^T (A^{-1})^T) \cdot (A^{-1}(\underline{x}-\underline{\mu}))] \\ &= (\underline{x}-\underline{\mu})^T ((A^{-1})^T A^{-1})(\underline{x}-\underline{\mu})\end{aligned}$$

und

$$\begin{aligned}((A^{-1})^T A^{-1})\Sigma &= [(A^{-1})^T A^{-1}][AA^T] = (A^{-1})^T (A^{-1}A)A^T \\ &= (A^{-1})^T A^T = (AA^{-1})^T = E^T = E,\end{aligned}$$

d.h. $(A^{-1})^T A^{-1} = \Sigma^{-1}$.

Die Behauptung folgt jetzt aus der Definition der Dichtefunktion.

Mehrdimensionale Normalverteilung

Erinnerung. Sei $\underline{X} \sim N(\underline{0}; E_n)$ standardnormalverteilt. Es gilt

$$f_{\underline{X}}(t_1, \ldots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{t_i^2}{2}} = \prod_{i=1}^{n} f_{X_i}(t_i),$$

wo f_{X_i} die Dichtefunktion von X_i ist. Also sind die Koordinaten eines standardnormalverteilten Zufallsvektors standardnormalverteilt und unabhängig.

Mehrdimensionale Normalverteilung

Behauptung. Sei $\underline{Y} = (Y_1, \dots, Y_n) \sim N(\underline{\mu}; \Sigma)$ ein nicht singulärer n-dimensionaler normalverteilter Zufallsvektor. Dann gilt $Y_i \sim (\mu_i; \Sigma_{i,i})$.

Beweis (Skizze). Sei

$$\underline{Y} = A\underline{X} + \underline{\mu},$$

wo $\mu \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, $\det A \neq 0$, und $\underline{X}$ standardnormalverteilt ist, dann folgt

$$Y_i = a_{i1}X_1 + \cdots + a_{in}X_n + \mu_i,$$

wo die Variablen $X_i \sim N(0; 1)$ gemeinsam unabhängig sind.

Mehrdimensionale Normalverteilung

Beweis (Skizze).

- Es ist leicht zu sehen, dass

$$\sum_{j=1}^{k-1} a_{ij} X_j \qquad \text{und} \qquad a_{ik} X_k$$

  zwei unabhängige Variablen für alle $1 < k \leq n$ sind.
- Durch vollständige Induktion beweist man, dass $\sum_{j=1}^{n} a_{ij} X_j$, und deshalb auch Y_i normalverteilt sind.
- Jeder Schritt der Induktion folgt aus der früher bewiesenen Behauptung, dass die Summe von zwei unabhängigen normalverteilten Variablen normalverteilt ist.

Mehrdimensionale Normalverteilung

Beweis (Skizze).

Die Werte der Parameter können aus $Y_i = a_{i1}X_1 + \cdots + a_{in}X_n + \mu_i$ anhand der linearität des Erwartungswertes und der Unabhängigkeit von $X_1, \ldots, X_n$ abgeleitet werden:

$$\mathbb{E}(Y_i) = \mu_i, \quad \mathbb{D}^2(Y_i) = \sum_{j=1}^{n} a_{ij}^2 \mathbb{D}^2(X_i) = \sum_{j=1}^{n} a_{ij}^2 = (AA^T)_{i,i} = \Sigma_{i,i}.$$

Mehrdimensionale Normalverteilung

Behauptung. Sei $(Y_1, Y_2)^T \sim N(\underline{\mu}; \Sigma)$ ein nicht singulärer zweidimensionaler normalverteilter Zufallsvektor.

a) Die (eindimensionale) Zufallsvariable $c_1 Y_1 + c_2 Y_2$ ist für alle $c_1, c_2 \in \mathbb{R}$ entweder normalverteilt, oder konstant.

b) Falls $\mathrm{corr}(Y_1, Y_2) = 0$ gilt, dann sind Y_1 und Y_2 unabhängig.

c) Die Regression $\mathbb{E}(Y_2 \mid Y_1)$ ist die lineare Regression von Y_2 auf Y_1, d.h.

$$\mathbb{E}(Y_2 \mid Y_1) = \frac{b}{a} Y_1 + \left(\mu_2 - \frac{b}{a} \mu_1 \right),$$

wo

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Mehrdimensionale Normalverteilung

Beweis. a) Es gibt eine 2×2-Matrix $A = (a_{ij})_{1 \leq i,j, \leq 2}$ und ein Vektor $(\mu_1, \mu_2)^T$, für die

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

gilt, wo X_1 und X_2 standardnormalverteilt und unabhängig sind. Deshalb

$$\begin{aligned} c_1 Y_1 + c_2 Y_2 &= c_1(a_{11}X_1 + a_{12}X_2 + \mu_1) + c_2(a_{21}X_1 + a_{22}X_2 + \mu_2) \\ &= (c_1 a_{11} + c_2 a_{21})X_1 + (c_1 a_{12} + c_2 a_{22})X_2 + c_1\mu_1 + c_2\mu_2. \end{aligned}$$

Wenn mindestens einer der Koeffizienten der ersten zwei Glieder Null ist, dann ist der Behauptung offenbar. Sonst sind aber diese Glieder unabhängig und normalverteilt, und dann ist auch ihre Summe normalverteilt.

Mehrdimensionale Normalverteilung

Beweis. b) Nehmen wir an, dass $\mathrm{corr}\,(Y_1, Y_2) = 0$, also $\mathrm{cov}\,(Y_1, Y_2) = 0$ gilt, dann hat die Konvarinzmatrix des Vektors $(Y_1, Y_2)^T$ die Form

$$\Sigma = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

$\underline{Y} = (Y_1, Y_2)^T$ ist nicht singulär, also existiert Σ^{-1}, und

$$\Sigma^{-1} = \frac{1}{\det \Sigma} \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix} = \frac{1}{ad} \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix} = \begin{pmatrix} a^{-1} & 0 \\ 0 & d^{-1} \end{pmatrix},$$

also ist die Dichtefunktion von $\underline{Y}$

$$f_{\underline{Y}}(s, t) = \frac{1}{2\pi\sqrt{ad}} e^{-\frac{(s-\mu_1)^2}{2a} - \frac{(t-\mu_1)^2}{2d}}.$$

Mehrdimensionale Normalverteilung

Beweis. Also sind die Randverteilungen durch die Dichtefunktionen

$$f_{Y_1}(s) = \frac{1}{\sqrt{2\pi a}} e^{-\frac{(s-\mu_1)^2}{2a}}, \qquad f_{Y_2}(t) = \frac{1}{\sqrt{2\pi d}} e^{-\frac{(t-\mu_2)^2}{2d}},$$

angegeben, folglich gilt $f_{\underline{Y}}(s,t) = f_{Y_1}(s) f_{Y_2}(t)$, also sind die Variablen Y_1 und Y_2 unabhängig.

Mehrdimensionale Normalverteilung

Wir beweisen Teil c) nicht.