Flerdimensionale fordelinger

Erik Michaelsen Nielsen

AALBORG UNIVERSITET

Masterprojekt

Institut for Matematiske Fag
Aalborg Universitet

Forår 2015

Forord

Dette masterprojekt er udarbejdet af Erik Michaelsen Nielsen på Aalborg Universitet, Institut for Matamatiske Fag, i perioden fra juni 2014 til marts 2015.

Projektet omhandler flerdimensionale fordelinger og består af 60 sider og indeholder ét appendiks. Oplagsantallet er 4.

Referencer er lavet efter Harvard-metoden og henvisninger i teksten vil se ud som følgende for bøger og tidsskrifter [efternavn, årstal]. Websider vil være henvist til på følgende måde [hovedsiden, årstal]. Projektet er i høj grad inspireret af bøgerne [Olofsson & Andersson, 2012] og [Seber, 1984].

Forfatteren vil gerne takke vejleder Bo Rosbjerg, der har været til stor hjælp med gode råd til den kontekstuelle og matematiske del af projektet.

Erik Michaelsen Nielsen

31-03-2015

Abstract

This projekt "Flerdimensionale fordelinger/Multivariate distributions", written by Erik Michaelsen Nielsen of Department of Mathematical Sciences at Aalborg University, deals with multivariate distributions.
A random variable is a function that associates a unique numerical value with every outcome of an experiment. However, in some experiments the outcomes can be complicated, and hence the use of only one random variable is not sufficient. For that reason, it becomes applicable to look at multidimensional random vectors and multivariate distributions. In addition the components of the random vectors can be examined simontaniusly, in order to detect their connection. By describing the random variables only one by one, we do not get all the possible information.
This projekt starts by describing two-dimentional random vectors. Basic operators and definitions regarding joint distributions are dealt with. Successively, several multivariate distributions, including the multivariate normal distribution and the Wishart distribution, are introduced and explained. Many of these distributions are applied in statistics. It is however beyond the scope of this projekt to decribe the applications of the multivariate distributions in that regard.

Indhold

Kapitel 1

Indledning

Sandsynlighedsteori er den del af matematiken, der omhandler beregninger af sandsynligheder for forskellige udfald af et eksperiment. Da det er et matematisk fundament for statistik, er sandsynlighedsteori nødvendigt i mange aktiviteter, som involverer en kvantitativ analyse af data.

Ideen til sandsynlighedsteori kom oprindeligt fra observationer eller udfald, som var knyttet til spil. Man mente, at en mønt var uærlig, hvis der ikke var et forhold tæt på $\frac{1}{2}$ mellem det totale antal gange det var blevet plat og det totale antal kast, efter at mønten havde været kastet n gange, hvor n er et stor tal. Ligeledes blev en terning opfattet som uærlig, hvis der ikke, efter et stort antal kast, var slået ca. $\frac{1}{6}$ toere.

Betragtninger af denne type gav anledning til, at man tolkede sandsynligheden for et specifikt udfald af et eksperiment, som forholdet mellem antallet af succesfulde udfald og totale antal udfald. Dermed kunne sandsynligheden udregnes som forholdet, i det lange løb, mellem antallet af gunstige udfald og mulige udfald.

Det er dog ikke altid muligt at udregne antallet af mulige udfald, hvilket gør denne metode uanvendelig. Desuden har flere eksperimenter ikke numeriske udfald. Derfor har man indført stokastiske variable for bedre at kunne behandle og beskrive eksperimenter. De stokastiske variable, med deres forskellige fordelinger, medvirker til, at man kan udregne sandsynligheder for et givent udfald af et eksperiment. Dog kan udfaldene være mange og komplekse, og derfor er det ikke altid tilstrækkeligt kun at bruge én stokastisk variabel. Dette giver anledning til at kigge på flerdimentionale stokastiske variable og flerdimentionale fordelinger.

Komponenterne i flerdimensionale stokastiske variable kan desuden undersøges samtidigt, således at det er sammenhængen mellem dem, der bliver belyst. Beskrives observationerne fra et eksperiment ud fra flere éndimensionale stokastiske variable, bliver denne sammenhæng ikke afdækket.

Dette projekt har netop til formål at beskrive flerdimensionale stokastiske variable og deres fordelinger. Der tages udgangspunkt i todimensionale stokastiske variable, hvor elementære operatorer og definitioner for simultane fordelinger bearbejdes. Herefter videreføres begreberne til flerdimensionale fordelinger, hvor flere konkrete fordelinger introduceres. Flere af de nævnte fordelinger finder anvendelse indenfor statistik, men det ligger imidlertid udover dette projektets mål, at beskrive disse anvendelser af flerdimensionale fordelinger.

Det forudsættes, at læseren har et indgående kendskab til elementær sandsynlighedsregning, endimensionale stokastiske variable og fordelinger.

Kapitel 2

Todimensionale fordelinger

Stokastiske variable bliver introduceret for at beskrive observationer af eksperimenter med tilfældige udfald. Mange typer af eksperimenter kan beskrives af en endimensional stokastisk variabel, men der findes også eksperimenter, hvor udfaldene er mere komplekse. Her kan observationerne eksempelvis beskrives af en todimensional (X, Y) stokastisk variabel. Ved kun at undersøge de stokastiske variable en ad gangen, er der ofte en sammenhæng mellem de dem, der ikke bliver belyst. Derfor er det relevant at undersøge todimensionale stokastiske variable [Olofsson & Andersson, 2012].

2.1 Den simultane fordelingsfunktion

Definition 2.1.1. *Lad X og Y være stokastiske variable. Da er parret (X, Y) en todimensional stokastisk variabel.*

Ideen er nu at beskrive de todimensionale stokastiske variable ud fra fordelingsfunktioner, sandsynlighedsfunktioner og tæthedsfunktioner på tilsvarende måde som for de endimensionale stokastiske variable.

Definition 2.1.2. *Den simultane fordelingsfunktion for en todimensional stokastisk variabel (X,Y)er defineret som*

$$F(x,y) = P(X \leq x,\ \ Y \leq y), \qquad x,y \in \mathbb{R}.$$

Fordelingsfunktionen er en funktion af to reelle variable. Den har tilsvarende egenskaber som endimentionale fordelingsfunktioner, men da det kræver tre dimentioner at tegne grafen for funktionen, er den ikke altid så nem at visualisere. Desuden er de todimensionale fordelingsfunktioner heller ikke tilsvarende egnede til udregning af sandsynligheder. For endimensionale stokastiske variable er ethvert udfald på formen $X = x$ eller $a < X < b$, og de tilsvarende sandsynligheder kan udtrykkes ved fordelingsfunktionen som

$$P(X = x) = F(x) - \lim_{y \uparrow x} F(y)$$

$$P(a < X < b) = F(b) - F(a), \qquad a \leq b$$

Det bliver dog anderledes svært for todimensionale stokastiske variable. Fordelingsfunktionens værdi i et punkt (x,y) er sandsynligheden for at $X \leq x$ og $Y \leq y$, hvilket er det samme som sandsynligheden for at (X, Y) findes i det uendelige reaktangel beskrevet af mængden $]-\infty, x] \times]-\infty, y]$. For et endeligt rektangel $T =]a,b] \times]c,d]$ kan det vises, at

$$P(a < X \leq b, c < X \leq d) = F(b,d) - F(b,c) - F(a,d) + F(a,c).$$

Der findes dog mange andre typer af mængder i to dimensioner. Hvis fx C er enhedscirklen, $C = (x,y) : x^2 + y^2 \leq 1$, kan $P((X,Y) \in C)$ ikke udtrykkes direkte vha. fordelingsfunktionen. Andre typer af mængder i R^2 giver anledning til det samme problem.
Hvis fordelingsfunktionen F for (X,Y) er kendt, kan de marginale fordelingsfunktioner $F_X(x)$ og $F_Y(y)$ for (X,Y) findes. Da udfaldet $Y \leq y$ er det samme som udfaldet $X < \infty, Y \leq y$, hvilket ikke sætter nogen grænse for X, gælder, at udfladet har sandsynligheden

$$P(X < \infty, Y \leq y) = F(\infty,y),$$

heraf følger sætningen.

Sætning 2.1.1. *Hvis (X,Y) har fordelingsfunktionen F, har X og Y henholdsvis fordelingsfunktionerne*

$$F_X(x) = F(x,\infty) \quad \textit{og} \quad F_Y(y) = F(\infty, y),$$

hvor $x,y \in \mathbb{R}$.

2.2 Diskrete stokastiske variable

Ligesom for andre stokastiske variable, skelnes der for de todimensionale mellem diskrete og kontinuerte.

Definition 2.2.1. *Hvis X og Y er diskrete stokastiske variable, gælder at (X,Y) er en todimensional diskret stokastisk variabel.*

Hvis værdimængden af X er $\{x_1, x_2, \dots\}$ og værdimængden af Y er $\{y_1, y_2, \dots\}$, bliver værdimængden af (X,Y) til $\{(x_j, y_k),\ \ j,k = 1, 2, \dots\}$, hvilket også er en tællelig mængde. Derfor kan sandsynlighedsfunktionen defineres på tilsvarende måde som for endimensionale stokastiske varable.

Definition 2.2.2. *Hvis (X,Y) er en todimensional stokastisk variabel med værdimængde $\{(x_j,y_k),\ \ j,k = 1, 2, \dots\}$, gælder, at funktionen*

$$p(x_j,y_k) = P(X = x_j, Y = y_k)$$

er sandsynlighedsfunktion for (X,Y).

Eksempel 2.2.1. *To piller tages tilfældigt fra et pilleglas, der indeholder fem hovedpinepiller, to sedativer og seks kvalmestillende piller. X og Y er henholdsvis antallet af hovedpinepiller og antallet af sedativer, blandt de to piller, der tilfældigt udvælges fra pilleglasset. Den todimensionale stokastiske variabel (X,Y) kan nu antage følgende værdier:* $(0,0), (0,1), (1,0), (1,1), (0,2)$ *og* $(2,0)$. *For at finde sandsynligheden for udfaldet* $(1,0)$*, altså tilfældet hvor man får én af de fem hovedpinepiller, ingen af de to* *sedativer og én af de seks kvalmestillende piller, gøres følgende. Antallet af muligheder hvor dette valg kan forekomme er*

$$\binom{5}{1}\binom{2}{0}\binom{6}{1} = 30.$$

Det totale antal af muligheder pillerne kan vælges på er

$$\binom{13}{2} = 78.$$

Da alle pillerne vælges fuldstændigt tilfældigt, er alle 78 muligheder lige sandsynlige, derfor kan sandsynligheden for udfaldet (1,0) *udregnes til* $\frac{30}{78} = \frac{5}{13}$*. Tilsvarende kan syndsynligheden for udfaldet* (1,1) *udregnes til*

$$\frac{\binom{5}{1}\binom{2}{1}\binom{6}{0}}{78} = \frac{10}{78} = \frac{5}{39}.$$

Fortsættes på denne måde, opnår man værdierne i tabel 2.1.

Tabel 2.1: Sandsynligheder for de forskellige udfald af udvælgelsen

		j: 0	j: 1	j: 2
	0	$\frac{5}{26}$	$\frac{5}{13}$	$\frac{5}{39}$
k	1	$\frac{2}{13}$	$\frac{5}{39}$	0
	2	$\frac{1}{78}$	0	0

Sandsynlighedsfunktionen kan skrives som.

$$p(j,k) = \frac{\binom{5}{j}\binom{2}{k}\binom{6}{2-j-k}}{78}, \quad j = 0,1,2; \quad k = 0,1,2; \quad 0 \leq j+k \leq 2.$$

□

Hvis sandsynlighedsfunktionen for den todimensionale stokastiske variabel (X,Y) er kendt, kan den marginale sandsynlighedsfunktion af de endimensionale stokastiske variable X og Y findes [Econ,2004].

Sætning 2.2.1. *Hvis* (X,Y) *har sandsynlighedsfunktionen* p*, gælder, at de marginale sandsynlighedsfunktioner af* X *og* Y *kan findes som*

$$p_X(x_j) = \sum_{k=1}^{\infty} p(x_j, y_k), \qquad j = 1,2,\dots\ ,$$

$$p_Y(y_k) = \sum_{j=1}^{\infty} p(x_j, y_k), \qquad k = 1,2,\dots\ .$$

Da udfaldet $\{Y = y_k\}$ kan beskrives ud fra den todimensionale stokastiske variabel (X,Y) som $\{X = \text{"Hvad som helst"}, Y = y_k\}$, får man

$$p_Y(y_k) = \sum_{j=1}^{\infty} P(X = x_j, Y = y_k) = \sum_{j=1}^{\infty} p(x_j, y_k).$$

Den tilsvarende formel for P_X opnås på samme måde.

Eksempel 2.2.2. *De marginale sandsynlighedsfunktioner tilhørende eksempel 2.2.1, kan altså findes ud fra formlerne i sætning 2.2.1. Hvis man ønsker at finde sandsynlighedsfunktionen for X, tager man altså udgangspunkt i sandsynlighedsfunktionen for (X,Y)*

$$p(j,k) = \frac{\binom{5}{j}\binom{2}{k}\binom{6}{2-j-k}}{78}, \quad j = 0,1,2; \quad k = 0,1,2; \quad 0 \leq j+k \leq 2.$$

For et bestemt j varierer k fra 0 til 2, samtidig må summen af j og k ikke overstige 2, hvilket giver

$$p_X(j) = \sum_{k=0}^{j+k=2} \frac{\binom{5}{j}\binom{2}{k}\binom{6}{2-j-k}}{78} = p(j,0) + p(j,1) + p(j,2).$$

Da der ikke længere tages højde for, om pillen er sedativ eller kvalmestillende, kan udtrykket omskrives til

$$p_X(j) = \frac{\binom{5}{j}\binom{8}{2-j}}{\binom{13}{2}},$$

hvilket er en hypergeometrisk fordeling med parametrene 13, 5 og 2,
$X \sim hypergeom(13,5,2)$. Dermed kan middelværdien og variansen findes til

$$E[X] = \frac{2 \cdot 5}{13} = \frac{10}{13}, \qquad Var[X] = 2 \cdot \frac{13-2}{13-1} \cdot \frac{5}{13} \cdot (1 - \frac{5}{13}) = \frac{220}{507}.$$

□

2.3 Kontinuerte stokastiske variable

En todimensional stokastisk variabel (X,Y) kan også være kontinuert.

Definition 2.3.1. *En ikke negativ funktion $f(x,y)$ defineret over xy-planen kaldes en simultan tæthedsfunktion for de kontinuerte stokastiske variable X og Y, hvis og kun hvis*

$$P((X,Y) \in A) = \iint_A f(x,y)dx\,dy, \qquad \textit{for alle delmængder } A \subseteq \mathbb{R}^2$$

$$\textit{og} \qquad \iint_{\mathbb{R}^2} f(x,y)dx\,dy = 1.$$

Der integreres altså over et todimensionalt område A. Hvis valget af A fx er
$A = \{(u,v) : u \leq x, v \leq y\} =]-\infty, x] \times]\infty, y]$, får man

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)du\,dv,$$

hvilket giver anledning til den følgende sætning.

Sætning 2.3.1. *Hvis (X,Y) er en todimentinal kontinuert stokastisk variabel med simultan tæthedsfunktion f og simultan fordelingsfunktion F, gælder*

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y), \qquad x, y \in \mathbb{R}.$$

Tæthedsfunktionen $f(x,y)$ er altså et mål for, hvor sandsynligt det er at (X,Y) ligger i en omegn af punktet (x,y). Ligesom ved det endimensionale tilfælde er sandsynligheden 0 for, at (X,Y) er lig med (x,y) for alle x og y. Bemærk at $F(\infty,\infty) = 1$ følger af definition 2.3.1.

Eksempel 2.3.1. *Lad den todimensionale stokastiske variabel have tæthedsfunktionen*

$$f(x,y) = 2(2y - x), \qquad 0 \leq x \leq 1, \; 0 \leq y \leq 1.$$

Vi ønsker nu at bestemme den tilhørende simultane fordelingsfunktion $F(x,y)$.

$$F(x,y) = 2\int_0^y \int_0^x (2k - j)dj \; dk = 2\int_0^y (2kx - \frac{1}{2}x^2)dk = 2\left[k^2 x - \frac{1}{2}x^2 k\right]_{k=0}^{y} \quad \Rightarrow$$
$$F(x,y) = 2y^2 x - x^2 y, \qquad 0 \leq x \leq 1, \; 0 \leq y \leq 1.$$

□

Hvis (X,Y) er en todimentinal kontinuert stokastisk variabel, kaldes funktionerne $f_X(x)$ og $f_Y(y)$ de marginale tæthedsfunktioner. Det nedre index indikerer, at $f_X(x)$ er defineret ud fra den stokastiske variabel X. Den marginale tæthed er den tæthed, der opstår, når alt information omkring den stokastiske variabel Y ignoreres. Heraf følger nedenstående sætning.

Sætning 2.3.2. *Antag at (X,Y) er en todimentinal kontinuert stokastisk variabel med tæthedfunktion f. Da er X og Y kontinuerte stokastiske variable med marginale tæthedsfunktioner,*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad x \in \mathbb{R} \;,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \quad y \in \mathbb{R} \;.$$

Eksempel 2.3.2. *Vi ønsker at bestemme de marginale tæthedsfunktioner $f_Y(y)$ og $f_X(x)$ ud fra den simultane tæthedsfunktion i eksempel 2.3.1.*
Den simultane tæthedsfunktion er

$$f(x,y) = 2(2y - x), \qquad 0 \leq x \leq 1, \; 0 \leq y \leq 1,$$

Hvilket i følge Sætning 2.3.2 giver

$$f_Y(y) \;=\; 2\int_0^1 (2y - x)dx \;=\; 2\left[2xy - \frac{1}{2}x^2\right]_{x=0}^{1} \;=\; 4y - 1, \qquad 0 \leq y \leq 1 \;,$$

og

$$f_X(x) \;=\; 2\int_0^1 (2y - x)dy \;=\; 2\left[y^2 - xy\right]_{y=0}^{1} \;=\; 2 - 2x, \qquad 0 \leq x \leq 1 \;.$$

□

2.4 Betingede fordelinger

Ideen bag betinget sandsynlighed er, at bestemme sandsynligheden for et udfald indtræffer, givet at et andet udfald er indtruffet.

Definition 2.4.1. *Lad X og Y være diskrete stokastiske variable henholdsvis med værdimængderne $\{x_1,x_2,\dots\}$ og $\{y_1,y_2,\dots\}$ og simultan sandsynlighedfunktion p. Da er den betingede sandsynlighedsfunktion af X givet $Y = y_k$ defineret som*

$$p_X(x_j|y_k) = \frac{p(x_j,y_k)}{p_Y(y_k)} \qquad \textit{for alle } j \in \{1,2,\dots\} \ .$$

Antag at en begivenheden beskrives ud fra den diskrete stokastiske variabel X. Sandsynligheden af udfaldet $\{X = x\}$ givet $\{Y = y\}$ kan betragtes som værdien den betingede sandsynlighedsfunktion af X i x givet udfaldet$\{Y = y\}$. Hvis Y er en diskret stokastisk variabel, kan definition 2.4.1 bruges til udregning af sandsynligheden.

Eksempel 2.4.1. *Der kastes en terning samt en mønt. En indikatorvariabel introduceres til resultatet af møntkastet, således at hvis det bliver plat, er værdien af indikatorvariablen 1, ellers er den nul. To stokastiske variable defineres nu til at beskrive udfaldet:*

X : *antallet af øjne ved slag med terningen.*

Y : *summen af antallet af øjne ved slag med terningen og værdien af indikatorvariablen.*

Sandsynligheden for et bestemt udfald (x, y) er givet i tabel 2.2, fx er

Tabel 2.2: De simultane og marginale sandsynligheder af X og Y

		Y							
		1	2	3	4	5	6	7	
X	1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	0	$\frac{1}{6}$
	2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0	$\frac{1}{6}$
	3	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	$\frac{1}{6}$
	4	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	$\frac{1}{6}$
	5	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{6}$
	6	0	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
		$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	1

$P(X=5, Y=6) = \frac{1}{12}$. Hvis sandsynligheden for udfaldet $x = 5$ givet $y = 6$ ønskes bestemt, kan definition 2.4.1 nu bruges:

$$p(5|6) = \frac{p(6{,}5)}{p_Y(6)} = \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{2}$$

Altså er sandsynligheden for at terningen viser fem øjne $\frac{1}{2}$, hvis summen af øjne og indikatorvariblen er 6 [Econ,2004].

□

I det kontinuerte tilfælde får ideen om betinget sandsynlighed en noget anderledes betydning end i det diskrete tilfælde. Hvis X og Y begge er kontinuerte stokastiske variable, er $P(X = x|Y = y)$ ikke defineret, fordi sandsynligheden i ét punkt er identisk med nul.
Da vædien af Y er kendt, kan sandsynligheden for at X er mindre end en bestemt værdi dog stadig udregnes. Derfor giver det stadig mening at undersøge fordelingen af X givet $Y = y$. Der opskrives en ny definition, der efterligner det diskrete tilfælde.

Definition 2.4.2. *Lad (X, Y) være en todimensional kontinuert stokastisk variabel med simultan tæthedsfunktion $f(x, y)$. Da gælder, at den betingede tæthedsfunktion af X givet $Y = y$ er*

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad x \in \mathbb{R} .$$

Udtrykket fortolkes som tæthedsfunktionen af X, hvis man ved at $Y = y$. For at finde den tilhørende fordelingsfunktion integreres tæthedsfunktionen og man får

$$F_X(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_X(s|y)ds, \qquad x \in \mathbb{R} ,$$

eller mere generelt

$$P(X \in A|Y = y) = \int_A f_X(s|y)ds, \qquad A \subseteq \mathbb{R} .$$

Eksempel 2.4.2. *Lad den simultane tæthedsfunktion af den todimensionale kontinuerte stokastiske variabel (X, Y) være givet ved*

$$f(x,y) = \begin{cases} 6(x-y) & 0 \leq y \leq x \leq 1 , \\ 0 & \textit{ellers.} \end{cases}$$

I følge sætning 2.3.2 findes de marginale tæthedsfunktioner til

$$\begin{aligned} f_X(x) &= \int_0^x f(x,y)dy = \int_0^x 6(x-y)dy \\ &= 6\left[xy - \frac{1}{2}y^2\right]_{y=0}^{x} = 6x^2 - 3x^2 = 3x^2, \qquad 0 \leq y \leq x \leq 1 , \\ f_Y(y) &= \int_y^1 f(x,y)dx = \int_y^1 6(x-y)dx = 6\left[\frac{1}{2}x^2 - yx\right]_{x=y}^{1} \\ &= 6((\frac{1}{2} - y) - (\frac{1}{2}y^2 - y^2) = 3 - 6y + 3y^2, \qquad 0 \leq y \leq x \leq 1 . \end{aligned}$$

Nu kan de betingede tæthedsfunktioner af X givet $Y = y$ og Y givet $X = x$ skrives som

$$\begin{aligned} f_X(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{6(x-y)}{3-6y+3y^2} = \frac{2(x-y)}{1-2y+y^2}, \qquad 0 \leq y \leq x \leq 1 , \\ f_Y(y|x) &= \frac{f(x,y)}{f_X(x)} = \frac{6(x-y)}{3x^2} = \frac{2(x-y)}{x^2}, \qquad 0 \leq y \leq x \leq 1 . \end{aligned}$$

For $0 \leq y \leq x \leq 1$ ønsker man, at betemme sandsynligheden for at $Y \leq \frac{1}{2}$ givet at $x = \frac{4}{5}$. Først bestemmes tæthedsfunktionen af Y, når $x = \frac{4}{5}$ som

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2(x-y)}{x^2} = \frac{2(\frac{4}{5}-y)}{(\frac{4}{5})^2} = \frac{(\frac{8}{5}-2y)}{\frac{16}{25}} = \frac{5}{2} - \frac{25}{8}y.$$

Herefter findes sandsynligheden

$$P\left(Y \leq \frac{1}{2}|X = \frac{4}{5}\right) = \int_{-\infty}^{\frac{1}{2}} f_Y(y|x)dt = \int_0^{\frac{1}{2}} \frac{5}{2} - \frac{25}{8}ydy$$
$$= \left[\frac{5}{2}y - \frac{25}{16}y^2\right]_{y=0}^{\frac{1}{2}} = \frac{5}{4} - \frac{25}{64} = \frac{55}{64}.$$

□

Den følgende sætning er en kontinuert udgave af loven om total sandsynlighed.

Sætning 2.4.1. *Lad (X, Y) være en todimensional kontinuert stokastisk variabel, Da gælder*

$$(I) \qquad f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)f_Y(y)dy, \qquad x \in \mathbb{R}$$

$$(II) \qquad P(X \in A) = \int_{-\infty}^{\infty} P(X \in A|Y = y)f_Y(y)dy, \qquad A \subseteq \mathbb{R}$$

Bevis

Ifølge definition 2.4.2 gælder

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)} \Rightarrow f_X(x|y)f_Y(y) = f(x,y) \ .$$

Anvendes nu sætning 2.3.2 får man

$$\int_{-\infty}^{\infty} f_X(x|y)f_Y(y)dy = \int_{-\infty}^{\infty} f(x,y)dy = f_X(x) \ ,$$

hvilket besviser (I).
For at bevise (II) udnyttes at $P(X \in A|Y = y) = \int_A f_X(x|y)dx$ sammen med (I).

$$P(X \in A) = \int_A f_X(x)dx = \int_A \int_{-\infty}^{\infty} f_X(x|y)f_Y(y)dydx =$$
$$\int_{-\infty}^{\infty} \int_A f_X(x|y)f_Y(y)dxdy = \int_{-\infty}^{\infty} P(X \in A|Y = y)f_Y(y)dy,$$

hvilket fuldfører beviset. ■

For at finde $P(X \in A)$, vælges en bestemt værdi, $Y = y$, og $P(X \in A|Y = y)$ udregnes. Herefter udregnes et vægtet gennemsnit af alle mulige værdier af Y, hvor tæthedsfunktionen bruges til at udregne vægtene. Da X er kontinuert anvendes et integral til denne udregning. Hvis der eksempelvis gælder at $A =]-\infty, x]$, får man

$$F_X(x) = \int_{-\infty}^{\infty} F_X(x|y)f_X(x)dx.$$

Sætning 2.4.2 er en anden brugbar version af loven om total sandsynlighed, til udregning af sandsynligheder af begivenheder, hvor både X og Y indgår.

Sætning 2.4.2. *Lad (X,Y) være en todimensional kontinuert stokastisk variabel, Da gælder for $A \subseteq \mathbb{R}^2$*

$$P((X,Y) \in A) = \int_{-\infty}^{\infty} P((X,Y) \in A | Y = y) f_Y(y) dy,$$

hvor $P((X,Y) \in A | Y = y)$ udelukkende afhænger af udfaldet af X.

Eksempel 2.4.3. *Lad $Y \sim unif[1,3]$ og givet at $Y = y$ lad $X \sim unif[0,y^2]$. Man ønsker nu at bestemme den simultane tæthedsfunktion af (X,Y) samt sandsynligheden $P(X \leq \frac{Y}{2})$.*
Først findes tæthedsfunktionen af Y til

$$f_Y(y) = \frac{1}{2}, \qquad 1 \leq y \leq 3$$

Den betingede tæthedsfunktion af X givet at $Y = y$ kan nu findes til

$$f_X(x|y) = \frac{1}{y^2}, \qquad 0 \leq x \leq y^2$$

Den simultane tæthedsfunktion bliver ifølge definition 2.4.2

$$f(x,y) = f_X(x|y) f_Y(y) = \frac{1}{2y^2}, \qquad 1 \leq y \leq 3, \qquad 0 \leq x \leq y^2$$

Området A sættes til $\{(x,y) : 0 \leq x \leq \frac{y}{2}, 1 \leq y \leq 3\}$, og man får

$$P((X,Y) \in A | Y = y) = P(X \leq \frac{Y}{2} | Y = y) = \int_A f_X(x|y) dx = \int_0^{\frac{y}{2}} \frac{1}{y^2} dx = \frac{y}{2y^2} = \frac{1}{2y} .$$

Nu følger af sætning 2.4.2, at

$$P(X \leq \frac{Y}{2}) = P((X,Y) \in A) = \int_1^3 \frac{1}{2y} \frac{1}{2} dy = \left[\frac{1}{4} ln\ y\right]_{y=1}^{3} = \frac{1}{4} ln\ 3 .$$

□

2.5 Uafhængige stokastiske variable

To udfald A og B siges at være uafhængige, hvis sandsynligheden for at de begge indtræffer, er det samme som produktet af sandsynlighederne for hvert af udfaldene indtræffer, altså

$$P(A \cap B) = P(A)P(B).$$

Anvendes stokastiske variable X og Y for udfaldene A og B sammen med ovenstående, giver det følgende definition.

Definition 2.5.1. *To stokastiske variable siges at være uafhængige, hvis*

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B) \qquad \textit{for alle} \quad A, B \subseteq \mathbb{R} .$$

Hvilket vil sige at X og Y er uafhængige, hvis den simultane sandsynlighed kan skrives som produktet af de marginale sandsynligheder. Følgende sætning kan nu opstilles.

Sætning 2.5.1. *Lad X være en stokastisk variabel med fordelingsfunktion $F_X(x)$, og lad Y være en stokastisk variabel med fordelingsfunktion $F_Y(y)$. Lad desuden X og Y have simultan fordelingsfunktion $F(x,y)$. Da gælder at X og Y er uafhængige, hvis og kun hvis.*

$$F(x,y) = F_X(x) F_Y(y), \qquad \textit{for alle } x,y \in \mathbb{R} .$$

Sætning 2.5.1 kan dog ikke nemt bevises.
Hvis to stokastiske variable X og Y er diskrete, kan uafhængigheden beskrives ud fra sandsynlighedsfunktioner som i sætningen herunder.

Sætning 2.5.2. *Antag at (X,Y) er en todimensional stokastisk variabel med sandsynlighedsfunktion p. Da gælder at X og Y er uafhængige, hvis og kun hvis.*

$$p(x,y) = p_X(x)p_Y(y), \qquad \textit{for alle } x,y \in \mathbb{R} \ .$$

Bevis

Først antages at $p(x,y) = p_X(x)p_Y(y)$, og der vælges to delmængder af $\mathbb{R}$, nemlig A og B, da gælder

$$P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} p(x,y) = \sum_{x \in A} p_X(x) \sum_{y \in B} p_Y(y) = P(X \in A)P(Y \in B) \ ,$$

hvilket viser at X og Y er uafhængige.
Antages omvendt, at X og Y er uafhængige, får man

$$p(x,y) = P(X = x, Y = y) = P(X = x)P(Y = y) = p_X(x)p_Y(y),$$

hvilket fuldfører beviset. ■

For kontinuerte stokastiske variable gælder en tilsvarende sætning.

Sætning 2.5.3. *Lad (X,Y) være en todimensional kontinuert stokastisk variabel med simultan tæthedsfunktion f. Da gælder at X og Y er uafhængige, hvis og kun hvis*

$$f(x,y) = f_X(x)f_Y(y) \qquad \textit{for alle } x,y \in \mathbb{R}.$$

Eksempel 2.5.1. *Lad den simultane tæthedsfunktion for de kontinuerte stokastiske variable X og Y, $f(x,y)$ være givet ved*

$$f(x,y) = \begin{cases} \frac{1}{2}xe^{-y} & 0 \leq x \leq 2, \quad 0 \leq y \ , \\ 0 & \textit{ellers.} \end{cases}$$

De marginale tæthedsfunktioner kan nu findes til

$$\begin{aligned} f_Y(y) &= \int_0^2 f(x,y)dx = \int_0^2 \frac{1}{2}xe^{-y}dx = \left[\frac{1}{4}x^2e^{-y}\right]_{x=0}^2 = e^{-y} \ , \\ f_X(x) &= \int_0^\infty f(x,y)dy = \int_0^\infty \frac{1}{2}xe^{-y}dy = \left[-\frac{1}{2}xe^{-y}\right]_{y=0}^\infty = 0 - (-\frac{1}{2}x) = \frac{1}{2}x \ . \end{aligned}$$

Det ses nu nemt, at $f(x,y) = f_X(x)f_Y(y)$, og dermed er X og Y uafhængige. □

2.6 Middelværdien af en funktion af stokastiske variable

Det kan være brugbart at betragte funktioner af typen $g : \mathbb{R}^2 \to \mathbb{R}$. Funktionen g bruges til danne en stokastisk variabel $g(X,Y)$ ud fra en todimentional stokastisk variabel (X,Y). For at udregne sandsynligheder af typen $P(g(X,Y) \in A)$, hvor $A \subseteq \mathbb{R}$, skal de mulige udfald af (X,Y), der afbildes ind i A, først identificeres.

Eksempel 2.6.1. *På en roulette, med 37 mulige udfald, (0-36), spinnes kuglen to gange. Lad X betegne udfaldet af det første spin, og Y udfaldet af det andet. Man ønsker at udregne sandsynligheden for, at de to udfald er mindst ti numre fra hinanden. Sandsynligheden $P(10 \leq |X-Y|)$ skal altså udregnes. Der gælder for de to diskrete stokastiske variable, at $X \sim unif[0{,}36]$ og $Y \sim unif[0{,}36]$. Da de to udfald er uafhængige af hinanden, gælder der at*

$$\sum_{k=0}^{36} \sum_{j=0}^{36} p(x_j, y_k) = \sum_{k=0}^{36} \sum_{j=0}^{36} p_X(x_j) p_Y(y_k) = 1 \ .$$

Sandsynligheden kan nu udregnes til

$$P(10 \leq |X - Y|) = 2 \cdot \sum_{k=10}^{36} \sum_{j=0}^{k-10} p(x_j, y_k) = 2 \cdot \sum_{k=10}^{36} \sum_{j=0}^{k-10} p_X(x_j) p_Y(y_k) =$$

$$2 \cdot \sum_{k=10}^{36} \sum_{j=0}^{k-10} \frac{1}{37} \cdot \frac{1}{37} = \frac{756}{1369} = 0{,}552 \ .$$

□

Middelværdien $E[g(X,Y)]$ kan findes ved at gøre brug af den følgende sætning.

Sætning 2.6.1. *Lad (X,Y) være en todimensional stokastisk variabel med sandsynlighedsfunktion p, eller tæthedsfunktion f og lad g være en funktion, $g : \mathbb{R}^2 \to \mathbb{R}$. Da gælder*

$$E[g(X,Y)] = \begin{cases} \sum\limits_{j=1}^{\infty} \sum\limits_{k=1}^{\infty} g(x_j, y_k) p(x_j, y_k), & \text{hvis } (X,Y) \text{er diskret,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy, & \text{hvis } (X,Y) \text{er kontinuert.} \end{cases}$$

Eksempel 2.6.2. *Den forventede forskel, på de to numre fra eksempel 2.6.1, kan nu udregnes ved at definere en funktion g*

$$g(x_j, y_k) = |x_j - y_k|.$$

Middelværdien af g, $E[g(X,Y)]$, kan nu findes ved at anvende sætning 2.6.1

$$E[g(X,Y)] = \sum_{j=0}^{36} \sum_{k=0}^{36} g(x_j, y_k) p(x_j, y_k) = \sum_{j=0}^{36} \sum_{k=0}^{36} \frac{|x_j - y_k|}{37^2} =$$

$$\frac{1}{37^2} \sum_{j=0}^{36} \sum_{k=0}^{36} |x_j - y_k| \ .$$

Udtrykkes udregnes ved brug af computer til

$$\frac{16872}{37^2} = 12{,}32 \ .$$

□

2.7 Middelværdi og varians af en sum

Der gælder, at middelværdien af en sum er det samme som summen af middelværdierne, jævnfør de to følgende sætninger.

Sætning 2.7.1. *Lad X og Y være stokastiske variable. Da gælder*

$$E[X+Y]=E[X]+E[Y]\ .$$

Bevis

Beviset udføres for det kontinuerte tilfælde. Først defineres en funktion $g(x,y)=x+y$, herefter anvendes sætning 2.6.1, og man får

$$E[X+Y]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy\ .$$

Anvendens sætning 2.3.2, får man

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy &= \int_{-\infty}^{\infty}xf_X(x)dx+\int_{-\infty}^{\infty}yf_Y(y)dy\\ &= E[X]+E[Y].\end{aligned}$$

■

Sætning 2.7.2. *Lad X og Y være stokastiske variable og lad a og b være reelle tal. Da gælder*

$$E[aX+bY]=aE[X]+bE[Y]$$

Bevis

Beviset udføres for det kontinuerte tilfælde. Først defineres en funktion $g(x,y)=ax+by$, herefter anvendes sætning 2.6.1, og man får

$$\begin{aligned}E[aX+bY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f(x,y)dxdy\\ &= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy+b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy\ .\end{aligned}$$

Anvendens sætning 2.3.2, får man

$$\begin{aligned}a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy+b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy &= a\int_{-\infty}^{\infty}xf_X(x)dx+b\int_{-\infty}^{\infty}yf_Y(y)dy\\ &= aE[X]+bE[Y].\end{aligned}$$

■

Det kan være nærliggende at tro, at der gælder tilsvarende for variansen, altså at $Var[X+Y]=Var[X]+Var[Y]$, men som eksempel 2.7.1 viser, gælder dette ikke altid for variansen.

Eksempel 2.7.1. *Hvis Y $unif[0,1]$ og $X=-Y+1$, gælder der at*

$$Var[X]=Var[Y]=\frac{1}{12}\Rightarrow Var[X]+Var[Y]=\frac{1}{6}\ .$$

Dog gælder der samtidigt at $Var[X+Y]=0$, da $X+Y=1$, og derfor at

$$Var[X+Y]\neq Var[X]+Var[Y]$$

□

Problemet som eksempel 2.7.1 illustrerer er, at selvom X og Y varierer individuelt, er der ingen variation i summen af de stokastiske variable. Det fremgår tydeligt, at de stokastiske variable i eksempel 2.7.1 er afhængige, og faktisk viser det sig, at uafhængighed er en forudsætning for at kunne addere varianser.

Sætning 2.7.3. *Lad X og Y være uafhængige stokastiske variablel. Da gælder*

$$\begin{aligned} (I) \qquad & E[XY] = E[X]E[Y] \ , \\ (II) \qquad & Var[X+Y] = Var[X] + Var[Y] \ . \end{aligned}$$

Bevis

(I) kan bevises ved at bruge sætning 2.6.1 og definere funktionen $g(xy) = xy$, man får

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy \ .$$

Her anvendes sætning 2.5.1, da de to stokastiske variable er uafhængige, og man får

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E[X]E[Y] \ .$$

(II) bevises ved at udnytte, at $Var[X] = E[X^2] - (E[X])^2$, heraf får man

$$\begin{aligned} Var[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= E[X^2] + 2E[XY] + E[Y^2] - (E[X])^2 - 2(E[X]E[Y]) - (E[Y])^2 \ . \end{aligned}$$

Nu benyttes (I), og man får

$$Var[X+Y] = E[X^2] + E[Y^2] - (E[X])^2 - (E[Y])^2 = Var[X] + Var[Y] \ .$$

■

Af sætning 2.7.1 fremgår det at middelværdioperatoren er lineær. Det samme er ikke tilfældet for variansoperatoren, men den følgende sætning er gældende.

Sætning 2.7.4. *Lad X være en stokatisk variabel, og lad a og b være reele tal, da gælder*

$$Var[aX+b] = a^2 Var[X] \ .$$

Bevis
Ud fra definitionen af varinasen kan følgende opstilles

$$\begin{aligned} Var[aX+b] &= E[(aX+b-E[aX+b])^2] \\ &= E[(aX+b-aE[X]-b)^2] = E[a^2(X-E[X])^2] \\ &= a^2 E[(X-E[X])^2] = a^2 Var[X] \ . \end{aligned}$$

■

Af sætning 2.7.4 og sætning 2.7.3 (II) kan følgende sætning udledes.

Sætning 2.7.5. *Lad X og Y være uafhængige stokastiske variable, og lad a og b være relle tal, da gælder*

$$Var[aX+bY] = a^2 Var[X] + b^2 Var[Y] \ .$$

2.8 Kovarians og korrelation

Indtil nu er de stokastiske variable blevet inddelt i katagorierne, afhængige eller uafhængige. Dette er en meget overordnet adskillelse og i tilfældet, hvor stokastiske variable er afhængige, vil man gerne kunne afgøre, i hvor høj grad de stokastiske variable afhænger af hinanden. Strategien der den, at man opstiller et mål for, hvor meget og hvordan de stokastiske variable afhænger af hinanden.

Kovarians

Definition 2.8.1. *Kovariansen af de stokastiske variable X og Y er defineret som*

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])]$$

Hvis større værdier af X giver større værdier af Y, gælder at $Cov[X,Y] > 0$, hvis større værdier af X giver mindre værdier af Y, gælder at $Cov[X,Y] < 0$.
Jo højere den numeriske værdi af $Cov[X,Y]$ er, jo højre grad af afhængighed er der mellem de to stokastiske variable. Udover de ovenstående sammenhænge gælder der også den følgende formel for kovariansen.

Sætning 2.8.1. *Lad X og Y være stokastiske variable, da gælder, at*

$$Cov[X,Y] = E[XY] - E[X]E[Y] \ .$$

Bevis

Sætningen bevises ved at anvende sætning 2.7.1.

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[Y]E[X] - E[Y]E[X] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

■

Kovariansen har følgende egenskab:
Hvis X og Y er uafhængige gælder der at $Cov[X,Y] = 0$.
Det modsatte er dog ikke altid tilfældet, som eksemplet her viser.

Eksempel 2.8.1. *Lad Y $unif[-1,1]$ og $X = Y^2$. Det er helt klart at X og Y er afhængige. Nu ønskes $Cov[X,Y]$ beregnet:*

$$\begin{aligned} Cov[X,Y] &= E[XY] - E[X]E[Y] \\ &= E[Y^3] - E[X]E[Y] \\ &= 0 - E[X] \cdot 0 \\ &= 0 \end{aligned}$$

□

Eksempel 2.8.2. *Lad X og Y være kontinurte stokastiske variable og lad deres simultane tæthedsfunktion været givet ved*

$$f(x,y) = \begin{cases} \frac{3}{8}x & \textit{hvis } 0 \leq Y \leq X \leq 2 \ , \\ 0 & \textit{ellers.} \end{cases}$$

Man ønsker at bestemme $Cov[X,Y]$, men først skal $E[XY]$ bestemmes:

$$\begin{aligned} E[XY] &= \int_0^2 \int_y^2 \frac{3}{8}x^2 y dx dy = \int_0^2 \left[\frac{1}{8}x^3 y\right]_{x=y}^2 dy \\ &= \int_0^2 (y - \frac{1}{8}y^4) dy = \left[\frac{1}{2}y^2 - \frac{1}{40}y^5\right]_{y=0}^2 \\ &= \frac{4}{2} - \frac{32}{40} = \frac{48}{40} = \frac{6}{5} \end{aligned}$$

Herefter udregnes middelværdien af X:

$$\begin{aligned} E[X] &= \int_0^2 \int_0^x \frac{3}{8}x^2 dy dx = \int_0^2 \left[\frac{3}{8}x^2 y\right]_{y=0}^x dx \\ &= \int_0^2 \frac{3}{8}x^3 dx = \left[\frac{3}{32}x^4\right]_{x=0}^2 \\ &= \frac{48}{32} = \frac{3}{2} \end{aligned}$$

Ligeledes udregnes middelværdien af Y:

$$\begin{aligned} E[Y] &= \int_0^2 \int_y^2 \frac{3}{8}xy dx dy = \int_0^2 \left[\frac{3}{16}x^2 y\right]_{x=y}^2 dy \\ &= \int_0^2 \frac{12}{16}y - \frac{3}{16}y^3 dy = \left[\frac{12}{32}y^2 - \frac{3}{64}y^4\right]_{y=0}^2 \\ &= \frac{48}{32} - \frac{48}{64} = \frac{48}{68} = \frac{3}{4} \end{aligned}$$

Kovariansen kan efterfølgende udregnes til

$$\begin{aligned} Cov[X,Y] = E[XY] - E[X]E[Y] &= \frac{6}{5} - \frac{3}{2}\frac{3}{4} \\ &= \frac{48}{40} - \frac{45}{40} = \frac{3}{40} \ . \end{aligned}$$

□

En anden vigtig egenskab ved kovariansen er, at den giver muligheden for at opstille en formel til udregning af variansen af en sum af stokastiske variable.

Sætning 2.8.2. *Lad X og Y være stokastiske variable, da gælder, at*

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X,Y] \ .$$

Bevis

Ved at anvende $Var[Z] = E[Z - E[Z]]^2$ får man

$$Var[X + Y] = E[(X + Y)^2] - (E[X + Y])^2 \ .$$

Her udnyttes sætning 2.7.1, hvilket giver

$$E[(X + Y)^2] - (E[X + Y])^2 = E[(X - E[X] + Y - E[Y])^2]$$

$$= E[(X - E[X])^2 + (Y - E[Y])^2 + 2(X - E[X])(Y - E[Y])] \ .$$

Anvendes definition 2.8.1 samt $Var[Z] = E[Z - E[Z]]^2$ igen, får man

$$E[(X-E[X])^2+(Y-E[Y])^2+2(X-E[X])(Y-E[Y])] = Var[X]+Var[Y]+2Cov[X,Y] \ .$$

■

Det ses at sætning 2.7.3 (II) er et specielt tilfælde af sætning 2.8.2, hvor $Cov[X,Y] = 0$.

Eksempel 2.8.3. *Med sætning 2.8.2 kan variansen af summen af de stokastiske variable fra eksempel 2.7.1, hvor $Y \sim unif[0,1]$ og $X = -Y+1$ og $Var[X] = Var[Y] = \frac{1}{12}$, beregnes. Kovariansen kan ifølge sætning 2.8.1 findes til*

$$Cov[X,Y] = Cov[-Y+1,Y] = E[(-Y+1)Y] - E[-Y+1]E[Y]$$
$$= E[-Y^2+Y] - (-E[Y]+1)E[Y] = -E[Y^2] + E[Y] + (E[Y])^2 - E[Y]$$
$$= -Var[Y] \ .$$

Bruges sætning 2.8.2, får man

$$Var[X+Y] = \frac{1}{12} + \frac{1}{12} + 2(-\frac{1}{12}) = 0 \ .$$

□

Flere vigtige egenskaber, for kovariansen, er opstillet i nedenstående sætning.

Sætning 2.8.3. *Lad X, Y og Z være stokastiske variable og lad a og b være reele tal, da gælder at*

$$\begin{array}{ll} (I) & Cov[aX,bY] = abCov[X,Y] \ , \\ (II) & Cov[X+Y,Z] = Cov[X,Y] + Cov[Y,Z] \ , \\ (III) & Cov[X,X] = Var[X] \ . \end{array}$$

Bevis

$$\begin{aligned} (I) \qquad Cov[aX,bY] &= E[abXY] - E[aX]E[bY] \\ &= abE[XY] - abE[X]E[Y] = ab(E[XY] - E[X]E[Y]) \\ &= abCov[X,Y] \end{aligned}$$

$$\begin{aligned} (II) \qquad Cov[X+Y,Z] &= E[(X+Y)Z] - E[X+Y]E[Z] \\ &= E[XZ+YZ] - (E[X]+E[Y])E[Z] \\ &= (E[XZ] - E[X]E[Z]) + (E[YZ] - E[Y]E[Z]) \\ &= Cov[X,Z] + Cov[Y,Z] \end{aligned}$$

$$\begin{aligned} (III) \qquad Cov[X,X] &= E[(X-E[X])(X-E[X])] = E[(X-E[X])^2] \\ &= Var[X] \end{aligned}$$

■

Kovariansen er et mål for afhængigheden mellem to stokastiske variable, men den har nogle begrænsninger. I udregningen af kovariansen tages nemlig ikke højde for, hvilke enheder der måles i. Faktisk kan en ændring i måleenheder føre til en ændring i kovariansen, som eksemplet her viser.

Eksempel 2.8.4. *Lad X være længden af den startbane, målt i meter, en flyvemaskine af typen Cherokee skal bruge, for at kunne lette, og lad Y være højden, hvorfra flyet letter, målt i fod. Graden af afhængighed mellem X og Y er nu $Cov[X,Y]$.*
Hvis man nu i stedet ønsker at måle startbanens længde i kilometer, betegnet K, og højden i meter, betegnet M, får man, $K = 0.001X$ og $M = 0{,}3048Y$, hvilket ifølge sætning 2.8.3 (I) giver

$$Cov[K,M] = Cov[0{,}001X; 0{,}3048Y] = 0{,}0003048Cov[X,Y].$$

Udregningerne viser at afhængigheden er væsentlig mindre end før, hvilket selvfølgelig ikke giver god mening. □

Eksempel 2.8.4 tydeliggør, at der er behov for at mål for afhængigheden mellem stokatiske variable, der ikke afhænger af måleenhder.

Korrelation

Korrelationskoefficienten mellem to stokastiske variable er et mål for interaktionen imellem dem. Den har samtidig den egenskab at være uafhængig af måleenheder, se eksempel 2.8.5, samt at være begrænset til værdier mellem -1 og 1.

Definition 2.8.2. *Korrelationskoefficienten af de stokastiske variable X og Y er defineret som*

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} .$$

Eksempel 2.8.5. *Man ønsker at udregne $\rho(K,M)$ for de stokastiske variable i eksempel 2.8.4*

$$\begin{aligned}\rho(K,M) &= \rho(0{,}001X; 0{,}3048Y)\\ &= \frac{Cov[0{,}001X, 0{,}3048Y]}{\sqrt{Var[0{,}001X]Var[0{,}3048Y]}} .\end{aligned}$$

Hvis sætning 2.8.3 (I) samt sætning 2.7.4 anvendes, får man

$$\begin{aligned}&= \frac{0{,}0003048Cov[X,Y]}{\sqrt{0{,}001^2Var[X]0{,}3048^2Var[Y]}}\\ &= \frac{0{,}0003048Cov[X,Y]}{0{,}0003048\sqrt{Var[X]Var[Y]}} = \rho(X,Y) .\end{aligned}$$

Udregningerne viser altså at $\rho(K,M) = \rho(X,Y)$. □

Værdierne 1 og -1 angiver den størst mulige grad af afhængighed mellem to stokastiske variable. Ligesom det var tilfældet for kovariansen, gælder desuden at korrelationskoefficienten er 0, hvis de to stokastiske variable er uafhængige.

Eksempel 2.8.6. *Ønsker man at udregne korrelationskoefficienten for de stokastiske variable i eksempel 2.8.2, skal* $E[X^2]$ *og* $E[Y^2]$ *først findes:*

$$\begin{aligned}
E[X^2] &= \int_0^2 \int_0^x \frac{3}{8}x^3 dydx \\
&= \int_0^2 \left[\frac{3}{8}x^3 y\right]_{y=0}^x dx = \int_0^2 \frac{3}{8}x^4 dx \\
&= \left[\frac{3}{40}x^5\right]_{x=0}^2 = \frac{96}{40} = \frac{12}{5} \\
E[Y^2] &= \int_0^2 \int_y^2 \frac{3}{8}xy^2 dxdy \\
&= \int_0^2 \left[\frac{3}{16}x^2 y^2\right]_{x=y}^2 dy = \int_0^2 \frac{12}{16}y^2 - \frac{3}{16}y^4 dy \\
&= \left[\frac{12}{48}y^3 - \frac{3}{80}y^5\right]_{y=0}^2 = \frac{96}{48} - \frac{96}{80} = \frac{4}{5}
\end{aligned}$$

Herefter kan $Var[X]$ *og* $Var[Y]$ *beregnes:*

$$\begin{aligned}
Var[X] &= E[X^2] - (E[X])^2 = \frac{12}{5} - (\frac{3}{2})^2 = \frac{3}{20} \\
Var[Y] &= E[Y^2] - (E[Y])^2 = \frac{4}{5} - (\frac{3}{4})^2 = \frac{19}{80}
\end{aligned}$$

Nu kan korrelationskoefficienten udregnes til

$$\begin{aligned}
\rho(X,Y) &= \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} \\
&= \frac{\frac{3}{40}}{\sqrt{\frac{3}{20} \cdot \frac{19}{80}}} = \frac{1}{19}\sqrt{57} \\
&= 0{,}397 \ .
\end{aligned}$$

□

Eksempel 2.8.6 viser altså, at der er en positiv korrelation mellem X og Y.

Kapitel 3

Flerdimensionale fordelinger

I forrige kapitel blev todimensionale stokastiske variable introduceret for at beskrive udfaldet af eksperimenter. I nogle tilfælde kan udfaldet af et eksperiment dog være så komplekst, at en todimensional stokastisk variabel ikke er tilstrækkelig. Dette giver anledning til at indføre flerdimensionale stokastiske variable og flerdimensionale fordelinger [Olofsson & Andersson, 2012].

3.1 Flerdimensionale stokastiske variable

Meget af det, der blev forklaret for todimensionale stokastiske variable i kapitel 2, kan på tilsvarende måde vises for d-dimensionale stokastiske variable. Der gælder altså tilsvarende, at $(X_1, \cdots, X_d)$ kaldes en d-dimensional stokastisk variabel, hvis $X_1, \cdots, X_d$ er endimensionelle stokastiske variable. Stokatiske variable, hvor dimensionen er større end to, vil fremover betegnes stokastiske vektorer.

Definition 3.1.1. *Lad $X = (X_1, X_2, \cdots, X_d)$ være en en d-dimensional stokastisk vektor. Da er middelværdivektoren $E[X]$ defineret som*

$$E[X] = (E[X_1], E[X_2], \cdots, E[X_d]).$$

Definition 3.1.2. *Lad $X = (X_1, X_2, \cdots, X_d)$ være en en d-dimensional stokastisk vektor, og antag, at enhver X_k er diskret. Da er den simultane sandsynlighedsfunktion defineret som*

$$p((x_1, \cdots, x_d) \in A) = P(X_1 = x_1, \cdots, X_d = x_d), \quad \textit{for alle } A \in \mathbb{R}^d \ .$$

Definition 3.1.3. *Hvis der findes en ikke negativ funktion $f : \mathbb{R}^d \to \mathbb{R}$, sådan at*

$$P((X_1, \cdots, X_d) \in A) = \int \cdots \int_A f(x_1, \cdots, x_d) dx_1 \cdots dx_d, \quad \textit{for alle } A \in \mathbb{R}^d$$

$$\textit{og} \qquad \int \cdots \int_{\mathbb{R}^2} f(x_1, \cdots, x_d) dx_1 \cdots dx_d = 1 \ ,$$

siges den stokastiske variabel at være kontinuert med simultan tæthedsfunktion f. Den tilhørende fordelingsfunktion, F, er defineret som

$$F(x_1, \cdots, x_d) = P((X_1 \leq x_1, \cdots, X_d \leq x_d) \ ,$$

og der er følgende sammenhæng mellem tæthedsfunktionen f og fordelingsfunktionen F

$$F(x_1, \cdots, x_d) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f(s_1, \cdots, s_d) ds_1 \cdots ds_d \ ,$$
$$f(x_1, \cdots, x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F(x_1, \cdots, x_d) \ .$$

De marginale funktioner kan, ligesom i det todimensionale tilfælde, findes ved at summere over den simultane sandsynlighedsfunktionen, eller ved at integrere over den simultane tæthedsfunktionen.
Hvis de stokastiske variable $X_1, X_2, \cdots$ kan antages at være gentagne målinger på den samme hændelse, vil de alle have den samme fordeling, og de vil desuden være uafhængige.

Eksempel 3.1.1. *I Yellowstone nationalpark ligger en gejser ved navn "Old Faithful". Den har navnet, da den siges at være i udbrud en gang i timen i gennemsnit. Dog påstår de fleste turister, at de skal vente mere end en time, for at se udbruddet, hvilket indikerer, at der er længere mellem udbrudene. Det antages at ventetiderne mellem hvert udbrud er uafhængige og ens fordelt, samt at de er* $unif \sim [30{,}90]$. *Man ønsker nu at udregne den gennemsnitlige ventetid for en turist, der ankommer til gejseren på et tilfældigt tidspunkt.*

Lad T_1,$T_2, \cdots$, *betegne tiderne mellem hvert udbrud, og lad* $\mu = E[T]$ *være den gennemsnitlige tid mellem to udbrud. Lad* T' *betegne tiden til næste udbrud fra et tilfældigt ankomststidspunkt t. Det gælder nu, at*

$$E[T'] = \frac{E[T^2]}{2\mu} = \frac{30^2 + 30 * 90 + 90^2}{3 \cdot (30 + 90)} min = \frac{65}{2} min = 32{,}5 min$$

Formlen viser, at den forventede længde af tidsintervallet, der indeholder tidspunktet t, er $\frac{E[T^2]}{\mu} = 65min$, *hvilket tydeligvis er længere end* $E[T] = (\frac{30+90}{2})min = 60min$. *Afvigelsen skyldes, at der er større sandsynlighed for at ankomme i et tidsinterval, der er længere end 60min end et der er mindre end 60min. Det er altså helt normalt at skulle vente længere end* $\frac{E[T]}{2} = 30min$ *på næste udbrud.* □

I flere situationer er det praktisk at ordne et observationssæt efter størrelse. Derfor indføres denne definition for stokastiske variable.

Definition 3.1.4. *Lad* $X_1, \cdots X_n$ *være uafhængige, og lad* $X_{(j)}$ *være den j'te mindste af* X_k*erne. De stokastiske variable* $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ *siges nu at være ordensvariable af de stokastiske variable* $X_1, \cdots X_n$.

Eksempel 3.1.2. *Lad* $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq X_{(5)} \leq X_{(6)}$ *være de ordens variable, der er forbundet med 6 uafhængige og ens fordelte observationer med tæthedsfunktionen*

$$f(x) = \frac{1}{2} x, \qquad hvor \quad 0 < x < 2 \ .$$

Man ønsker nu at finde sandsynligheden for at $X_{(4)} < 1$, *dvs.* $P(X_{(4)} < 1)$.

Den eneste måde $X_{(4)}$ *kan være mindre end 1, er hvis værdierne af mindst 4 af de stokastiske variable* X_1,X_2,X_3, X_4, X_5 *og* X_6 *er mindre end 1. Sandsynligheden for at et givent udfald er mindre end 1, kan udregnes til*

$$P(X_i \leq 1) = \frac{1}{2}\int_0^2 x dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_{x=0}^2 = \frac{1}{2}\left(\frac{1}{2} - 0\right) = \frac{1}{4}\ .$$

Hvis man kalder udfaldet $X_i < 1, i = 1, 2, \dots, 6$ for en succes, og man lader Z betegne antallet af successer i seks uafhængige forsøg, er Z en binomial fordelt stokastisk variabel med parametrene $n = 6$ og $p = 0.25$. For, at finde sandsynligheden for at $X_{(4)}$ er mindre end 1, skal følgende binomiale udregning foretages.

$$P(X_{(4)} < 1) = P(Z = 4) + P(Z = 5) + P(Z = 6)$$

$$= \binom{6}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^2 + \binom{6}{5}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^1 + \binom{6}{6}\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^0 = 0.0376$$

□

3.2 Middelværdi- og kovariansoperatorer

Betragtes n d-dimensionale stokastiske vektorer, kan det være anvendeligt at lade dem indgå i en matrix, $\mathbf{X} = [X_{ij}]$, dvs.

$$\mathbf{X} = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1d} \\ X_{21} & X_{22} & \cdots & X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nd} \end{pmatrix}\ ,$$

hvor hver række i matricen er en d-dimensional stokastisk vektor. Dette kaldes fremadrettet for en stokastisk matrix. Søjlevektorerne i $\mathbf{X}$ vil fremadrettet betegnes som $X_{j'}$, hvor $j = 1, \cdots, d$.
Begrebet middelværdi kan udvides til matricer [Seber, 1984].

Definition 3.2.1. *Lad $\boldsymbol{X}$ være den stokastiske matrix ovenfor. Da er middelværdi operatoren E, for matricen defineret som*

$$E[\boldsymbol{X}] = [(E[X_{ij}])] = \begin{pmatrix} E[X_{11}] & E[X_{12}] & \cdots & E[X_{1d}] \\ E[X_{21}] & E[X_{22}] & \cdots & E[X_{2d}] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_{n1}] & E[X_{n2}] & \cdots & E[X_{nd}] \end{pmatrix}.$$

Matricen,$E[\boldsymbol{X}]$, kaldes middelværdimatricen af $\boldsymbol{X}$.

Desuden får man ud fra linariteten af middelværdien

$$E[\mathbf{AXB} + \mathbf{C}\ ] = \mathbf{A}E[\mathbf{X}\ ]\mathbf{B} + \mathbf{C}\ .$$

Kovariansen af to stokastiske vektorer resulterer i en matrix.

Definition 3.2.2. *Lad X_1 og X_2 være to stokastiske vektorer. Da er kovariansmatricen, af de to stokastiske vektorer defineret som*

$$Cov[X_1, X_2] = [(Cov[X_{1i}, X_{2j}])] = [E[(X_{1i} - E[X_{1i}])(X_{2j} - E[X_{2j}])]] =$$
$$\begin{pmatrix} E[(X_{11} - E[X_{11}])(X_{21} - E[X_{21}]) & \cdots & E[(X_{11} - E[X_{11}])(X_{2c} - E[X_{2c}]) \\ \vdots & \ddots & \vdots \\ E[(X_{1d} - E[X_{1d}])(X_{21} - E[X_{21}]) & \cdots & E[(X_{1d} - E[X_{1d}])(X_{2c} - E[X_{2c}]) \end{pmatrix}.$$

Det ses af definition 3.2.2, at to stokastiske variable ikke nødvendigvis skal have samme dimension, for at man kan finde kovariansmatricen.

Sætning 3.2.1. *Lad X_1 og X_2 være to stokastiske vektorer. Da er deres kovariansmatix*

$$Cov[X_1, X_2] = E[X_1 X_2^T] - E[X_1]E[X_2]^T$$

Bevis

$$\begin{aligned}
Cov[X_1, X_2] &= [(Cov[X_1,X_2])] \\
&= [E[(X_1 - E[X_1])(X_2 - E[X_2])]] \\
&= E[(X_1 - E[X_1])(X_2 - E[X_2])^T] \\
&= E[X_1 X_2^T - X_1 E[X_2]^T - E[X_1]E[X_2]^T + E[X_1]E[X_2]^T] \\
&= E[X_1 X_2^T] - E[X_1]E[X_2]^T - E[X_1]E[X_2]^T + E[X_1]E[X_2]^T \\
&= E[X_1 X_2^T] - E[X_1]E[X_2]^T
\end{aligned}$$

■

To stokastiske vektorer siges at være uafhængige, hvis to vilkårlige komponenter, fra hver sin vektor, er uafhængige. Uafhængighed mellem to stokastiske vektorer stiller således ikke krav til uafhængighed mellem to komponenter fra samme vektor. Der gælder desuden at $Cov[X_1, X_2] = \mathbf{O}$, hvis X_1 og X_2 er uafhængige, hvilket ses af følgende omskrivning.

$$Cov[X_1, X_2] = E[X_1 X_2^T] - E[X_1]E[X_2]^T = E[X_1]E[X_2]^T - E[X_1]E[X_2]^T = \mathbf{O}$$

Sætning 3.2.2. *Lad X_1 og X_2 være stokastiske vektorer og lad $\boldsymbol{A}$ og $\boldsymbol{B}$ være matricer, da gælder*

$$Cov[\boldsymbol{A}X_1, \boldsymbol{B}X_2] = \boldsymbol{A}Cov[X_1, X_2]\boldsymbol{B}^T \ .$$

Bevis

$$\begin{aligned}
Cov[\mathbf{A}X_1, \mathbf{B}X_2] &= E[(\mathbf{A}X_1 - \mathbf{A}E[X_1])(\mathbf{B}X_2 - \mathbf{B}E[X_2])^T] \\
&= E[\mathbf{A}(X_1 - E[X_1])(X_2 - E[X_2])^T \mathbf{B}^T] \\
&= \mathbf{A}\ E[(X_1 - E[X_1])(X_2 - E[X_2])^T]\ \mathbf{B}^T \\
&= \mathbf{A}Cov[X_1, X_2]\mathbf{B}^T
\end{aligned}$$

■

Definition 3.2.3. *Lad $X = (X_1, X_2, \cdots, X_d)$ være en d-dimensional stokastiske vektor. Da er variansmatricen $Var[X]$ defineret som*

$$Var[X] = Cov[X, X] = [(Cov[X_i,X_j])] = [E[(X_i - E[X_i])(X_j - E[X_j])]] =$$

$$\begin{pmatrix} E[(X_1 - E[X_1])(X_1 - E[X_1]) & \cdots & E[(X_1 - E[X_1])(X_d - E[X_d]) \\ \vdots & \ddots & \vdots \\ E[(X_d - E[X_d])(X_1 - E[X_1]) & \cdots & E[(X_d - E[X_d])(X_d - E[X_d]) \end{pmatrix} \Rightarrow$$

$$Var[X] = \begin{pmatrix} Var[X_1] & Cov[X_1,X_2] & \cdots & Cov[X_1,X_d] \\ Cov[X_2,X_1] & Var[X_2] & \cdots & Cov[X_1,X_d] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_d,X_1] & Cov[X_d,X_2] & \cdots & Var[X_d] \end{pmatrix}$$

Det ses af definition 3.2.3, at variansmatricen for en d-dimensional stokastisk variabel X er en symmetrisk matix med variansen af komponenterne i X som diagonal. Af sætning 3.2.2 får man desuden

$$Var[\mathbf{A}X] = Cov[\mathbf{A}X,\mathbf{A}X] = \mathbf{A}Cov[X,X]\mathbf{A}^T = \mathbf{A}Var[X]\mathbf{A}^T \ .$$

Fremover bruges forkortelsen $\mathbf{\Sigma}$ til variansmatricer.

Eksempel 3.2.1. *I en have står 5 birketræer. For hvert træ måles, i meter, højden, omkreds af stammen samt aftanden fra jorden til nederste gren. Måledata er indsat i matrix* $\boldsymbol{X}_{obs}$.

$$\boldsymbol{X}_{obs} = \begin{pmatrix} 3{,}90 & 0{,}55 & 1{,}35 \\ 4{,}20 & 0{,}75 & 1{,}55 \\ 2{,}00 & 0{,}35 & 0{,}65 \\ 4{,}30 & 0{,}70 & 1{,}25 \\ 6{,}90 & 0{,}95 & 2{,}05 \end{pmatrix}$$

Hver rækkevektor x_i *er altså måledata for en observation af en tredimensional stokastisk variabel. Estimatatet af middelværdivektoren kan findes til*

$$\overline{x} = (\overline{x}_{1'}, \overline{x}_{2'}, \overline{x}_{3'}) = (4{,}26; 0{,}66; 1{,}37).$$

Værdierne i estimatet $\boldsymbol{S}$ *af variansmatricen* $\mathbf{\Sigma}$ *kan findes som*

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_{j'})(x_{ik} - \overline{x}_{k'}) \ ,$$

og man får

$$\boldsymbol{S} = \begin{pmatrix} 3{,}053 & 0{,}376 & 0{,}854 \\ 0{,}376 & 0{,}040 & 0{,}109 \\ 0{,}854 & 0{,}109 & 0{,}257 \end{pmatrix} \ .$$

Her er 3,053 den empiriske varians af højden på træerne, 0,040 er den empiriske varians af stammens omkreds og 0,257 er den empiriske varians af afstanden fra jorden til den nederste gren. Ligeledes kan det aflæses, at den empiriske kovarians mellem højden og omkredsen er 0,376, den empiriske kovarians mellem højden og afstanden er 0,854, og den empiriske kovarians mellem omkredsen og afstanden er 0,109.

Definition 3.2.4. *Antag at* $X_1, X_2, \cdots, X_n$ *er n uafhængige d-dimensionale stokastiske vektorer med henholdsvis middelværdivektorer* $E[X_1], E[X_2], \cdots, E[X_n]$ *og variansmatricer* $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \cdots, \mathbf{\Sigma}_n$. *Da defineres den generalisrede kvadratiske form som*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j^T = \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} \ ,$$

hvor $\boldsymbol{X} = (X_1 \cdots X_n)^T$ *og* $\boldsymbol{A} = [a_{ij}]$ *er en symmetrisk matrix.*

Sætning 3.2.3. *Lad* $X_1, X_2, \cdots, X_n$ *være n uafhængige stokastiske vektorer med henholdsvis middelværdivektorer* $E[X_1], E[X_2], \cdots, E[X_n]$ *og variansmatricer* $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \cdots, \mathbf{\Sigma}_n$, *da gælder*

$$E[\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}] = \sum_{i=1}^{n} a_{ii}\mathbf{\Sigma}_i + E[\boldsymbol{X}^T]\boldsymbol{A}E[\boldsymbol{X}] \ .$$

Bevis

$$
\begin{aligned}
E[\mathbf{X}^T\mathbf{A}\mathbf{X}] &= E[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j^T] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E[a_{ij}X_iX_j^T] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E[(X_i - E[X_i])(X_j - E[X_j])^T \\
&\qquad + X_iE[X_j]^T + E[X_i]X_j^T - E[X_i]E[X_j]^T] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(E[(X_i - E[X_i])(X_j - E[X_j])^T] \\
&\qquad + E[X_i]E[X_j]^T + E[X_i]E[X_j]^T - E[X_i]E[X_j]^T \\
&= \sum_{i=1}^{n} a_{ij}Cov[X_i, X_j] + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E[X_i]E[X_j]^T \\
&= \sum_{i=1}^{n} a_{ii}\mathbf{\Sigma}_i + E[\mathbf{X}^T]\mathbf{A}E[\mathbf{X}]
\end{aligned}
$$

■

Der gælder desuden, hvis $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_n$, at

$$E[\mathbf{X}^T\mathbf{A}\mathbf{X}] = tr(\mathbf{A})\mathbf{\Sigma} + E[\mathbf{X}^T]\mathbf{A}E[\mathbf{X}] \ .$$

hvor $tr(\mathbf{A})$ er sporet af $\mathbf{A}$, og $\mathbf{\Sigma}$ er fælles variansmatrix til de n uafhængige stokastiske variable.

Et andet begreb, der anvendes for flerdimensionale fordelinger, er Kroneckerproductet af to matricer.

Definition 3.2.5. *Lad $\boldsymbol{A}$ og $\boldsymbol{B}$ henholdsvis være en $m \times m$ og en $n \times n$ matrix, da er*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1m}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2m}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mm}\boldsymbol{B} \end{pmatrix}$$

Kroneckerproductet af $\boldsymbol{A}$ og $\boldsymbol{B}$, hvor $\boldsymbol{A} \otimes \boldsymbol{B}$ er en $mn \times mn$ matrix.

Hvis man igen betragter n uafhængige d-dimensionale stokastiske vektorer $X_1, X_2, \cdots, X_n$ med fælles variansmatrix $\mathbf{\Sigma}$, gælder, at vektoren
$Y = (X_1^T X_2^T \cdots X_n^T)^T$ har variansmatrix

$$Var[Y] = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{\Sigma} \end{pmatrix} = \mathbf{I}_n \otimes \mathbf{\Sigma},$$

hvor $\mathbf{I}_n$ er $n \times n$ enhedsmaticen.

3.3 Momentfrembringende funktioner

Momentfrembringende funktioner introduceres, da de anvendes i flere beviser, bl.a. for den flerdimensionale normalfordeling.

Definition 3.3.1. *Lad $X = (X_1, \cdots, X_n)$ være en d-dimensional stokastisk variabel, da er*

$$M_X(t_1, t_2, \cdots, t_n) = M_X(t) = E[e^{t^T X}]$$

den momentfrembringende funktion for X.

Sætning 3.3.1. *Lad $X = (X_1, \cdots, X_n)$ være en d-dimensional stokastisk variabel, da gælder*

$$M_{\boldsymbol{A}X+b}(t) = e^{t^T b} M_X(\boldsymbol{A}^T t) \ .$$

Bevis

Sætningen vises ved følgende omskrivning af udtrykket for den momentfrembringende funktion, [People-1,2001]:

$$\begin{aligned} M_{\mathbf{A}X+b}(t) &= E[e^{t^T(\mathbf{A}X+b)}] \\ &= e^{t^T b} E[e^{t^T \mathbf{A}X}] \\ &= e^{t^T b} E[e^{(\mathbf{A}^T t)^T X}] \\ &= e^{t^T b} M_X(\mathbf{A}^T t) \end{aligned}$$

■

Sætning 3.3.2. *Lad to stokastiske variable X og Y være uafhængige og lad dem have samme dimension, da gælder*

$$M_{X+Y}(t) = M_X(t) M_Y(t) \ .$$

Bevis

Sætningen vises ved følgende omskrivning af udtrykket for den momentfrembringende funktion, [People-1,2001]:

$$\begin{aligned} M_{X+Y}(t) &= E[e^{t^T(X+Y)}] \\ &= E[e^{t^T(X)+t^T(Y)}] \\ &= E[e^{t^T(X)} e^{t^T(Y)}] \\ &= E[e^{t^T(X)}] \ \ E[e^{t^T(Y)}] \\ &= M_X(t) M_Y(t) \end{aligned}$$

■

Kapitel 4

Den flerdimensionale normalfordeling

Til mange endimensionale kontinuerte fordelinger svarer en tilsvarende flerdimensional fordeling med den egenskab, at de endimensionale marginalfordelinger alle tilhører samme type. Der findes bl.a. flerdimensionelle normalfordelinger, som dette afsnit vil belyse [Seber, 1984].

4.1 Definitioner

Der findes to definitioner af den flerdimentionale normalfordeling. Den første definerer fordelingen ud fra tæthedsfunnktionen, definition 4.1.1. Den anden definition er baseret på den unikke egenskab ved den flerdimensionale normalfordeling, at enhver linearkombination af dens komponenter er endimensionale normalfordelinger, definition 4.1.2.

Definition 4.1.1. *Lad $X = (X_1, X_2, \cdots, X_d)$ være en en d-dimensional stokastisk vektor. Da siges X at have en d-dimensional normalfordeling hvis, dens tæthedsfunktion er*

$$f(x) = (2\pi)^{-\frac{d}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\theta)^T \mathbf{\Sigma}^{-1}(x-\theta)}, \quad hvor$$

$-\infty < x_j < \infty,\ j = 1, 2, \cdots, d,\ E[X] = \theta \in \mathbb{R}^d$ *og* $Var[X] = \mathbf{\Sigma} \in \mathbb{R}^{d\times d}$.
Bemærk at $\mathbf{\Sigma}$ skal være en positiv definit matrix.

Når X har en d-dimensional normalfordeling skrives $X \sim N_d(\theta, \mathbf{\Sigma})$. Hvis alle X_j er uafhængige og har en endimensional normalfordeling $N(\theta_j, \sigma^2), j = 1, 2, \cdots, d$, ses ud af definition 4.1.1 at $X \sim N_d(\theta, \sigma^2 \mathbf{I}_d)$. Ligeledes gælder $X - \theta \sim N_d(0, \mathbf{\Sigma})$.

Definition 4.1.2. *Lad X være en d-dimensional stokastisk vektor, $E[X] = \theta$ og $Var[X] = \mathbf{\Sigma} > \boldsymbol{O}$ og lad $u^T X$ have endimensional normalfordeling for alle $u \neq 0$. Da siges X at have en flerdimensional normalfordeling $X \sim N_d(\theta, \mathbf{\Sigma})$.*

Den momentfrembringende funktion for den flerdimensionale normalfordeling, se sætning 4.1.1, kan bruges til at vise, at de to definitioner 4.1.1 og 4.1.2 er ækvivalente, se sætning 4.1.2.
Situationen hvor $\mathbf{\Sigma}$ er singulær, kan inkluderes i definition 4.1.2, hvis restriktionen ændres til $\mathbf{\Sigma} \geq \mathbf{O}$, så der findes mindst ét u, hvor $Var[u^T X] = 0$.

Sætning 4.1.1. *Antag at $X \sim N_d(\theta, \mathbf{\Sigma})$. Da findes den momentfrembringende funktion af X som*

$$M_X(t) = E[e^{(t^T X)}] = e^{(t^T \theta + \frac{1}{2} t^T \mathbf{\Sigma} t)} .$$

Bevis

Fra definitionen af momentfrembringende funktioner 3.3.1 får man $M_X(t_1, t_2, \cdots, t_d) = M_X(t)$, der kan skrives som

$$\begin{aligned} M_X(t) &= E[e^{t^T X}] \\ &= \int_{\mathbb{R}^d} e^{t^T x}(2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\theta)^T\mathbf{\Sigma}^{-1}(x-\theta)}d\Omega\ . \end{aligned}$$

Ved at bruge den lineære transformation $X = \mathbf{\Sigma}^{\frac{1}{2}}Y + \theta$ med Jacobideterminant $\frac{\partial(X_1, X_2, \cdots, X_d)}{\partial(Y_1, Y_2, \cdots, Y_d)} = |\mathbf{\Sigma}^{\frac{1}{2}}|$, bliver tæthedsfunktionen for $Y = (Y_1, Y_2, \cdots, Y_d)$

$$f(y) = (2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(y^T\mathbf{\Sigma}^{\frac{1}{2}})\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}^{\frac{1}{2}}y)}|\mathbf{\Sigma}^{\frac{1}{2}}|\ .$$

Ved at skifte variabel bliver udregningen følgende

$$\begin{aligned} M_X(t) &= \int_{\mathbb{R}^d} e^{t^T(\mathbf{\Sigma}^{\frac{1}{2}}x+\theta)}(2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}(x^T\mathbf{\Sigma}^{\frac{1}{2}})\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}^{\frac{1}{2}}x)}|\mathbf{\Sigma}^{\frac{1}{2}}|d\Omega \\ &= \int_{\mathbb{R}^d} e^{t^T(\mathbf{\Sigma}^{\frac{1}{2}}x+\theta)}(2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}x^Tx}d\Omega\ . \end{aligned}$$

Omskrivningen er mulig, da der gælder $|\mathbf{\Sigma}| = |\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}^{\frac{1}{2}}| = |\mathbf{\Sigma}^{\frac{1}{2}}|^2 \Rightarrow |\mathbf{\Sigma}|^{\frac{1}{2}} = |\mathbf{\Sigma}^{\frac{1}{2}}|$. Videre gælder

$$\begin{aligned} &= \int_{\mathbb{R}^d} e^{t^T(\mathbf{\Sigma}^{\frac{1}{2}}x+\theta)}(2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}x^Tx}d\Omega \\ &= e^{t^T\theta}\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}e^{t^T\mathbf{\Sigma}^{\frac{1}{2}}x-\frac{1}{2}x^Tx}d\Omega \\ &= e^{t^T\theta+\frac{1}{2}t^T\mathbf{\Sigma}t}\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}(x^Tx-2t^T\mathbf{\Sigma}^{\frac{1}{2}}x+t^T\mathbf{\Sigma}t)}d\Omega \\ &= e^{t^T\theta+\frac{1}{2}t^T\mathbf{\Sigma}t}\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}(x-\mathbf{\Sigma}^{\frac{1}{2}}t)^T(x-\mathbf{\Sigma}^{\frac{1}{2}}t)}d\Omega\ . \end{aligned}$$

Da funktionen under integralet er tæthedsfunktionen for fordelingnen $N_d(\mathbf{\Sigma}^{\frac{1}{2}}t, \mathbf{I}_d)$, kan udtrykket forenkles til

$$M_X(t) = e^{t^T\theta+\frac{1}{2}t^T\mathbf{\Sigma}t}$$

■

Sætning 4.1.2. *De to definitioner 4.1.1 og 4.1.2 er ækvivalente, når der ses bort fra det singulære tilfælde.*

Bevis

Lad $X \sim N_d(\theta, \mathbf{\Sigma})$ være defineret som i definition 4.1.1. Nu gælder ifølge sætning

4.1.1, at $M_X(t) = e^{\left(t^T\theta+\frac{1}{2}t^T\mathbf{\Sigma}t\right)}$. Desuden har man, for enhver vektor u, ifølge sætning 3.3.1, at

$$\begin{aligned} M_{u^TX}(t) = M_X(ut) &= e^{\left((ut)^T\theta+\frac{1}{2}(ut)^T\mathbf{\Sigma}(ut)\right)} \\ &= e^{\left(u^T\theta t+\frac{1}{2}u^T\mathbf{\Sigma}ut^2\right)} \ . \end{aligned}$$

Lad nu u og $X \sim N_d(\theta, \mathbf{\Sigma})$ være defineret som i definition 4.1.2. Da følger af sætningerne 4.1.1 og 3.3.1, at $M_{u^TX}(t) = e^{\left(u^T\theta t+\frac{1}{2}u^T\mathbf{\Sigma}ut^2\right)}$. Vælges $t = 1$ og skrives vektoren u som t, får man for enhver t, at

$$M_{t^TX}(1) = e^{\left(t^T\theta\cdot 1+\frac{1}{2}t^T\mathbf{\Sigma}t\cdot 1^2\right)} \ ,$$

hvilket viser, at

$$M_X(t) = e^{\left(t^T\theta+\frac{1}{2}t^T\mathbf{\Sigma}t\right)} \ .$$

Da de to momentfrembringende funktioner er identiske, er definitionerne 4.1.1 og 4.1.2 ækvivalente [People-2,2001]. ■

4.2 Vigtige egenskaber

De følgende sætninger i afsnittet beskriver nogle af de vigtigste egenskaber ved den flerdimensionale normalfordeling.

Sætning 4.2.1. *Hvis $X \sim N_d(\theta, \mathbf{\Sigma})$ og $\boldsymbol{C}$ er en $q \times d$ matrix med rang q, gælder at*

$$\boldsymbol{C}X + b \sim N_q(\boldsymbol{C}\,\theta + b, \boldsymbol{C}\,\mathbf{\Sigma}\boldsymbol{C}^T) \ .$$

Bevis

Først anvendes sætning 3.3.1

$$M_{\mathbf{A}X+b}(t) = e^{t^Tb}M_X(\mathbf{A}^Tt) \ ,$$

og man får

$$\begin{aligned} M_{\mathbf{C}X+b}(t) &= e^{t^Tb}M_X(\mathbf{C}^Tt) \\ &= e^{t^Tb} \ e^{\left((\mathbf{C}^Tt)^T\theta+\frac{1}{2}(\mathbf{C}^Tt)^T\mathbf{\Sigma}(\mathbf{C}^Tt)\right)} \\ &= e^{\left(t^T(\mathbf{C}\theta+b)+\frac{1}{2}t^T\mathbf{C}\mathbf{\Sigma}\mathbf{C}^T\right)} \ . \end{aligned}$$

Da man ud fra sætning A.0.1 ser, at $\mathbf{C\Sigma C}^T > \mathbf{O}$, gælder der nu ifølge sætning 4.1.1 at

$$\mathbf{C}X + b \sim N_q(\mathbf{C}\theta + b, \mathbf{C\Sigma C}^T),$$

hvilket viser sætningen. Specielt gælder $\mathbf{C}X \sim N_q(\mathbf{C}\theta, C\mathbf{\Sigma C}^T)$. ■

Sætning 4.2.2. *Antag at $X \sim N_d(\theta, \mathbf{\Sigma})$ og lad*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \qquad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

hvor X_1 og θ_1 er d_1-dimensionale vektorer og $\mathbf{\Sigma}_{11}$ er en $d_1 \times d_1$ matrix og $d = d_1 + d_2$. Da gælder $X_1 \sim N_{d_1}(\theta_1,\mathbf{\Sigma}_{11})$.

Bevis

Ved at definere $\mathbf{C} = (\mathbf{I}_{d_1} \;\; \mathbf{O})$, der har rang d_1, og udnytte sætning 4.2.1 får man

$$(\mathbf{I}_{d_1} \;\; \mathbf{O})X \sim N_{d_1}\left((\mathbf{I}_{d_1} \;\; \mathbf{O})\theta \ , \quad (\mathbf{I}_{d_1} \;\; \mathbf{O})\mathbf{\Sigma}\begin{pmatrix}\mathbf{I}_{d_1} \\ \mathbf{O}\end{pmatrix}\right) \Rightarrow X_1 \sim N_{d_1}(\theta_1, \mathbf{\Sigma}_{11}) \ .$$

■

Ved at omarrangere komponterne i X kan enhver delmængde af komponenter X_j fra X, skrives som X_1. Derfor gælder, at en vilkårlig delmængde af komponenter fra X danner en stokastisk vektor, som er flerdimensional normalfordelt. Specielt vil ethver komponent være éndimensionalt normalfordelt.
Sætning 4.2.2 viser altså, at alle marginalfordelinger af en normalfordelt stokastisk vektor er normalfordelinger. Det omvendte er dog ikke gældende. I det følgende eksempel vises, at to stokastiske variable kan være normalfordelt uden at have en simultan normalfordeling.

Eksempel 4.2.1. *Lad $Y \sim N(0{,}1)$, og lad X være uafhængig af Y, sådan at $P(X = 1) = \frac{1}{2}$ og $P(X = -1) = \frac{1}{2}$. Nu defineres $Z = XY$, og man får*

$$\begin{aligned} P(Z \leq y) &= \frac{1}{2}P(Y \leq y) + \frac{1}{2}P(-Y \leq y) \\ &= \frac{1}{2}F(y) + \frac{1}{2}(1 - F(-y)) = F(y) \end{aligned}$$

Dermed er $Z \sim N(0{,}1)$, hvilket vil sige at både Z og Y er standard normalfordelt. Dog gælder der samtidigt at

$$P(Y + Z = 0) = P(X = -1) = \frac{1}{2}$$

Heraf ses, at $Y + Z$ ikke er normalfordelt og derfor er (Y,Z), jf. definition 4.1.2, heller ikke normalfordelt [Gut, 2009]. □

Sætning 4.2.3. *To normalfordelte stokastiske vektorer X_1 og X_2 er uafhængige, hvis og kun hvis $Cov[X_1, X_2] = \boldsymbol{O}$*

Bevis

Lad $X = \begin{pmatrix}X_1 \\ X_2\end{pmatrix}$ og $t = \begin{pmatrix}t_1 \\ t_2\end{pmatrix}$ sådan at X_2 og t_2 er d_1-dimensionale og X_1 og t_1 er d_2-dimensionale, samt at henholdvis X_1, X_2 og t_1, t_2 kan sammensættes til X og t. Nu gælder ifølge sætning 4.1.1, at

$$\begin{aligned} M_X(t) &= e^{(t^T\theta + \frac{1}{2}t^T\mathbf{\Sigma}t)} \\ &= e^{\left((t_1^T t_2^T)\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix} + \frac{1}{2}(t_1^T t_2^T)\begin{pmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{pmatrix}\begin{pmatrix}t_1 \\ t_2\end{pmatrix}\right)} \\ &= e^{\left(t_1^T\theta_1 + t_2^T\theta_2 + \frac{1}{2}t_1^T\mathbf{\Sigma}_{11}t_1 + \frac{1}{2}t_2^T\mathbf{\Sigma}_{21}t_1 + \frac{1}{2}t_1^T\mathbf{\Sigma}_{12}t_2 + \frac{1}{2}t_2^T\mathbf{\Sigma}_{22}t_2\right)} \\ &= e^{\left(t_1^T\theta_1 + \frac{1}{2}t_1^T\mathbf{\Sigma}_{11}t_1\right)} \; e^{\left(t_1^T\mathbf{\Sigma}_{12}t_2\right)} \; e^{\left(t_2^T\theta_2 + \frac{1}{2}t_2^T\mathbf{\Sigma}_{22}t_2\right)} \\ &= M_{X_1}(t_1) \; e^{\left(t_1^T\mathbf{\Sigma}_{12}t_2\right)} \; M_{X_2}(t_2) \\ &= M_{(X_1 \;\; 0)^T}(t) \; e^{\left(t_1^T\mathbf{\Sigma}_{12}t_2\right)} \; M_{(X_2 \;\; 0)^T}(t) \ . \end{aligned}$$

Da der gælder, at X_1 og X_2 er uafhængige, hvis og kun hvis $\begin{pmatrix} X_1 \\ 0 \end{pmatrix}$ og $\begin{pmatrix} 0 \\ X_2 \end{pmatrix}$ er uafhængige, får man ifølge sætning 3.3.2, at en nødvendig betingelse er

$$e^{\left(t_1^T \mathbf{\Sigma}_{12} t_2\right)} = 1 \ \text{ for alle t} \Leftrightarrow \mathbf{\Sigma}_{12} = \mathbf{O} \Leftrightarrow Cov[X_1, X_2] = \mathbf{O},$$

hvilket beviser at $Cov[X_1, X_2] = \mathbf{O}$ når X_1 og X_2 er uafhængige.
Antages omvendt at $Cov[X_1, X_2] = \mathbf{O}$, får man variansmatricen $\mathbf{\Sigma}$ for $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ til at være

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

Hvor $\mathbf{\Sigma}_{11}$ og $\mathbf{\Sigma}_{22}$ er variansmatricerne for X_1 og X_2. Heraf følger

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma}_{22}^{-1} \end{pmatrix}.$$

Nu ses at

$$\begin{aligned}
(x-\theta)^T \mathbf{\Sigma}^{-1} (x-\theta) &= \begin{pmatrix} x_1 - \theta_1 \\ x_2 - \theta_2 \end{pmatrix}^T \begin{pmatrix} \mathbf{\Sigma}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{\Sigma}_{22}^{-1} \end{pmatrix} \begin{pmatrix} x_1 - \theta_1 \\ x_2 - \theta_2 \end{pmatrix} \\
&= (x_1 - \theta_1)^T \mathbf{\Sigma}_{11}^{-1} (x_1 - \theta_1) + (x_2 - \theta_2)^T \mathbf{\Sigma}_{22}^{-1} (x_2 - \theta_2) \\
&= h_1(x_1) + h_2(x_2)\ .
\end{aligned}$$

Tæthedsfunktionen for X bliver

$$\begin{aligned}
& f_X(x_{11}, \cdots, x_{1d_1}, x_{21}, \cdots, x_{2d_2}) \\
&= (2\pi)^{-\frac{d_1+d_2}{2}} (|\mathbf{\Sigma}_{11}||\mathbf{\Sigma}_{22}|)^{-\frac{1}{2}} e^{-\frac{1}{2}(h_1(x_1)+h_2(x_2))} \\
&= (2\pi)^{-\frac{d_1}{2}} (|\mathbf{\Sigma}_{11}|)^{-\frac{1}{2}} e^{-\frac{1}{2}(h_1(x_1))} (2\pi)^{-\frac{d_2}{2}} (|\mathbf{\Sigma}_{22}|)^{-\frac{1}{2}} e^{-\frac{1}{2}(h_2(x_2))} \\
&= f_{X_1}(x_{11}, \cdots, x_{1d_1}) f_{X_2}(x_{21}, \cdots, x_{2d_2}),
\end{aligned}$$

hvilket viser at X_1 og X_2 er uafhængige og dermed fuldfører beviset. ∎

Sætning 4.2.4. *Hvis $V_i = \mathbf{A}_i X$, $i = 1, 2, \cdots, m$ og X er defineret som i sætning 4.2.2, da er alle V_i parvis uafhængige, hvis og kun hvis $Cov[V_i, V_j] = \mathbf{O}$ for alle $i \neq j$*

Bevis

Antag at $\mathbf{A}_i, i = 1, \cdots, m$, har fuld rang og $\sum_{i=1}^{m} rang(A_i) \leq d$ og lad

$$\begin{pmatrix} V_1 \\ \vdots \\ V_m \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 X \\ \vdots \\ \mathbf{A}_m X \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{pmatrix} X.$$

Betragt $V = \begin{pmatrix} V_i \\ V_j \end{pmatrix}, i \neq j$.

$$Cov[V_i, V_j] = \mathbf{O} \quad \Leftrightarrow \quad Cov[\mathbf{A}_i X, \mathbf{A}_j X] = \mathbf{O}.$$

Det følger af sætning 3.2.2, at

$$Cov[\mathbf{A}_i X, \mathbf{A}_j X] = \mathbf{O} \quad \Leftrightarrow \quad \mathbf{A}_i \mathbf{\Sigma} \mathbf{A}_j^T = \mathbf{O},$$

hvilket viser at rækkerne i $\mathbf{A}_i$ er ortogonale på rækkerne i $\mathbf{A}_j$ mht. det indre produkt $\langle a,b \rangle = a^T \mathbf{\Sigma} b$. Derfor har $\mathbf{A} = \begin{pmatrix} \mathbf{A}_i \\ \mathbf{A}_j \end{pmatrix}$ fuld rang, dvs.

rang $\begin{pmatrix}\mathbf{A}_i\\ \mathbf{A}_j\end{pmatrix} = \text{rang}(\mathbf{A}_i) + \text{rang}(\mathbf{A}_j) = p \leq d$. Dermed er $V \sim N_p(\mathbf{A}\theta, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)$, og der gælder ifølge sætning 4.2.3, at

$$Cov[V_i, V_j] = \mathbf{O} \quad \Leftrightarrow \quad V_i \text{ og } V_j \text{ er uafhængige.}$$

■

Sætning 4.2.5. *Hvis X er defineret som i sætning 4.1.1, gælder*

$$(X-\theta)^T\mathbf{\Sigma}^{-1}(X-\theta) \sim \chi^2_d \ .$$

Bevis

Igen anvendes den lineære transformation $X = \mathbf{\Sigma}^{\frac{1}{2}}Y + \theta$, hvor tæthedsfunktionen for $Y = (Y_1, Y_2, \cdots, Y_d)$ er

$$f(y) = (2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}(y^T\mathbf{\Sigma}^{\frac{1}{2}})\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}^{\frac{1}{2}}y)\right)}|\mathbf{\Sigma}^{\frac{1}{2}}| \ .$$

Udtrykket kan omskrives til

$$\begin{aligned} f(y) &= (2\pi)^{-\frac{d}{2}} e^{\left(-\frac{1}{2}(y^Ty)\right)} \\ &= (2\pi)^{-\frac{d}{2}} e^{\left(-\frac{1}{2}(y^Ty)\right)} \\ &= (2\pi)^{-\frac{d}{2}} e^{\left(-\frac{1}{2}(\sum_{i=1}^{d} y_i^2)\right)} \\ &= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \quad e^{\left(-\dfrac{y_i^2}{2}\right)}. \end{aligned}$$

Hvilket viser, at $Y_i \sim N(0{,}1), \ i = 1, 2, \cdots, d$ og $Y_1, Y_2, \cdots, Y_d$ er uafhængige. Man får

$$\begin{aligned} (X-\theta)^T\mathbf{\Sigma}^{-1}(X-\theta) &= (\mathbf{\Sigma}^{\frac{1}{2}}Y + \theta - \theta)^T\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}^{\frac{1}{2}}Y + \theta - \theta) \\ &= Y^TY = \sum_{i=1}^{d} Y_i^2 \sim \chi^2_d \ . \end{aligned}$$

$(X-\theta)^T\mathbf{\Sigma}^{-1}(X-\theta)$ er altså chi-i-anden fordelt med d frihedsgrader. ■

Sætning 4.2.6. *Hvis $X \sim N_d(\theta, \mathbf{\Sigma})$, gælder at $X^T\mathbf{\Sigma}^{-1}X$ har en ikke-central chi-i-anden fordeling med d frihedsgrader og ikke-centralitets parameter $\delta = \theta\mathbf{\Sigma}^{-1}\theta$.*

Bevis

Lad X være skrevet som $X = \mathbf{\Sigma}^{\frac{1}{2}}Y$. Man får

$$Y = \mathbf{\Sigma}^{-\frac{1}{2}}X \ \sim \ N_d(\mathbf{\Sigma}^{-\frac{1}{2}}\theta, \ \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{\Sigma}^{-\frac{1}{2}}) = N_d(\mathbf{\Sigma}^{-\frac{1}{2}}\theta, \ \mathbf{I}_d) \ .$$

Udtrykket $X^T\mathbf{\Sigma}^{-1}X$ kan herefter omskrives til

$$X^T\mathbf{\Sigma}^{-1}X = \ Y^T\mathbf{\Sigma}^{\frac{1}{2}} \ \mathbf{\Sigma}^{-1} \ \mathbf{\Sigma}^{\frac{1}{2}}Y = \ Y^TY.$$

Da Y er normalfordelt med middelværdien $\mathbf{\Sigma}^{-\frac{1}{2}}\theta$ og varians $\mathbf{I}_d$, gælder at Y^TY er chi-i-anden fordelt med d frihedsgrader og $(\mathbf{\Sigma}^{-\frac{1}{2}}\theta)^T(\mathbf{\Sigma}^{-\frac{1}{2}}\theta)$ som ikke-centralitets parameter. Altså

$$X^T\mathbf{\Sigma}^{-1}X \sim \ \chi^2(d, \ (\mathbf{\Sigma}^{-\frac{1}{2}}\theta)^T(\mathbf{\Sigma}^{-\frac{1}{2}}\theta)) = \chi^2(d, \ \theta^T\mathbf{\Sigma}^{-1}\theta)$$

■

Sætning 4.2.7. *Lad*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{d_1+d_2}\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\right).$$

Da gælder, at den betingede fordeling af X_2 givet $X_1 = x_1$ er $N_{d_2}(\theta_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(x_1 - \theta_1),\ \mathbf{\Sigma}_{22\cdot 1})$, hvor $\mathbf{\Sigma}_{22\cdot 1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$.

Bevis

Først introduceres $Z = X_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}X_1$, der er normalfordelt, da Z er en linearkombination af komponenter fra X. Vektoren $\begin{pmatrix} X_1 \\ Z \end{pmatrix}$ kan skrives som

$$\begin{pmatrix} X_1 \\ Z \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{d_1} & 0 \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \mathbf{I}_{d_2} \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Variansmatricen findes til

$$\begin{aligned}
Var\begin{pmatrix} X_1 \\ Z \end{pmatrix} &= \begin{pmatrix} \mathbf{I}_{d_1} & 0 \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \mathbf{I}_{d_2} \end{pmatrix}\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{I}_{d_1} & -(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1})^T \\ 0 & \mathbf{I}_{d_2} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ -\mathbf{\Sigma}_{21} + \mathbf{\Sigma}_{21} & -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{I}_{d_1} & -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \\ 0 & \mathbf{I}_{d_2} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Sigma}_{11} & -\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{12} \\ 0 & -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}_{11} & 0 \\ 0 & -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{22} \end{pmatrix}.
\end{aligned}$$

Ifølge sætning 4.2.3 er X_1 og Z er uafhængige. Fra ovenstående variansmatrix ser man at $Var[Z] = -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{22\cdot 1}$.
Middelværdien af Z kan udregnes til

$$E[Z] = E[X_2] - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}E[X_1] = \theta_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\theta_1 \ .$$

Der gælder altså at $Z \sim N_{d_2}(\theta_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\theta_1,\ \ \mathbf{\Sigma}_{22\cdot 1})$.
Udtrykkes X_2 ud fra Z, får man

$$X_2 = Z + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}X_1 \ .$$

Betinget med $X_1 = x_1$, får man følgende udtryk for X_2

$$\begin{aligned}
X_2|x_1 &= Z|x_1 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x_1 \\
&= Z + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x_1 \ .
\end{aligned}$$

Middelværdien af $X_2|x_1$ findes til

$$\begin{aligned}
E[X_2|x_1] &= EZ + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x_1 \\
&= \theta_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\theta_1 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}x_1 \\
&= \theta_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(x_1 - \theta_1).
\end{aligned}$$

og variansen er $Var[X_2|x_1] = Var[Z] = \mathbf{\Sigma}_{22\cdot 1}$. Dermed gælder

$$X_2|x_1 \sim N_{d_2}(\theta_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(x_1 - \theta_1), \mathbf{\Sigma}_{22\cdot 1}).$$

■

Eksempel 4.2.2. *Antag at $Y \sim N(\theta, \boldsymbol{\Sigma})$, hvor $\theta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ og $\boldsymbol{\Sigma} = \begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix}$. Man ønsker at finde den betingede fordelingsfunktion af $Y_1 + Y_2$ givet at $Y_1 - Y_2 = 0$. Ved at introducere de stokastiske variable $Z_1 = Y_1 + Y_2$ og $Z_2 = Y_1 - Y_2$ kan problemet reduceres til at finde den betingede fordeling af Z_1 givet $Z_2 = 0$. Det ses at Z kan skrives som*

$$Z = \boldsymbol{B}Y, \quad hvor \;\; \boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Ifølge sætning 4.2.1 gælder at $Z \sim N(\boldsymbol{B}\theta, \;\; \boldsymbol{B\Sigma B}^T)$, hvilket vil sige

$$Z \sim N\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \; \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\right) = N\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \; \begin{pmatrix} 12 & 4 \\ 4 & 8 \end{pmatrix}\right).$$

Derfor bliver tæthedsfunktionen af Z

$$\begin{aligned}
f_Z(z) &= (2\pi)^{-\frac{2}{2}} \left|\begin{pmatrix} 12 & 4 \\ 4 & 8 \end{pmatrix}\right|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}\begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}^T \begin{pmatrix} 12 & 4 \\ 4 & 8 \end{pmatrix}^{-1} \begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}\right)} \\
&= \frac{1}{2\pi\sqrt{80}} \; e^{\left(-\frac{1}{2}\begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}^T \begin{pmatrix} \frac{1}{10} & -\frac{1}{20} \\ -\frac{1}{20} & \frac{3}{20} \end{pmatrix} \begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}\right)} \\
&= \frac{1}{2\pi\sqrt{80}} \; e^{\left(-\frac{1}{2}\begin{pmatrix} \frac{z_1 - 2}{10} - \frac{z_2}{20} & -\frac{z_1 - 2}{20} + \frac{3z_2}{20} \end{pmatrix}\begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}\right)} \\
&= \frac{1}{2\pi\sqrt{80}} \; e^{\left(\begin{pmatrix} \frac{4 + z_2 - 2z_1}{40} & \frac{2 - z_1 - 3z_2}{40} \end{pmatrix}\begin{pmatrix} z_1 - 2 \\ z_2 \end{pmatrix}\right)} \\
&= \frac{1}{2\pi\sqrt{80}} \; e^{\left(\frac{(4 + z_2 - 2z_1)(z_1 - 2) + (2 - z_1 - 3z_2)z_2}{40}\right)} \\
&= \frac{1}{2\pi\sqrt{80}} \; e^{\left(\frac{z_1 - 1}{5} - \frac{z_1^2}{20} - \frac{3z_2^2}{40}\right)}.
\end{aligned}$$

Man kan desuden se at $Z_2 \sim N(0{,}8)$, hvilket giver følgende tæthedsfunktion

$$f_{Z_2}(z_2) = \frac{1}{\sqrt{2\pi}\sqrt{8}} \; e^{\left(-\frac{1}{2}\;\frac{z_2^2}{3}\right)}.$$

Afslutningsvis får man altså

$$\begin{aligned}
f_{Z_1|z_2}(z_1) = \frac{f_{Z_1,Z_2}(z_1, 0)}{f_{Z_2}(0)} &= \frac{\frac{1}{2\pi\sqrt{80}} \; e^{\left(\frac{z_1 - 1}{5} - \frac{z_1^2}{20}\right)}}{\frac{1}{\sqrt{2\pi}\sqrt{8}} \; e^{(-\frac{1}{2} \;\cdot\; 0\;)}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{10}} \; e^{\left(-\frac{1}{2}\;\frac{(z_1 - 2)^2}{10}\right)}.
\end{aligned}$$

Dette er tæthedsfunktionen for en $N(2,10)$ fordeling. Det ses desuden at

$$\theta_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(z_2 - \theta_2) = 2 + 4 \cdot \frac{1}{8} \cdot (0-0) = \quad 2,$$

og

$$\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} = 12 - 4 \cdot \frac{1}{8} \cdot 4 = 12 - 2 = 10,$$

hvilket stemmer overens med sætning 4.2.7. □

Sætning 4.2.8. *Lad $Y_1, Y_2, \cdots, Y_n$ være uafhængige og have fordelingerne $N_d(\theta_i, \mathbf{\Sigma}_i)$, $i = 1, 2, \cdots, n$. Da har $\sum_{i=1}^{n} a_i Y_i$ fordelingen $N_d\left(\sum_{i=1}^{n} a_i\theta_i,\ \sum_{i=1}^{n} a_i^2 \mathbf{\Sigma}_i\right)$.*

Bevis

$\sum_{i=1}^{n} a_i Y_i$ har følgende momentfrembringende funktion

$$\begin{aligned} M_{\sum_{i=1}^{n} a_i Y_i}(t) &= \prod_{i=1}^{n} M_{a_i Y_i}(t) = \prod_{i=1}^{n} M_{Y_i}(a_i t) = \prod_{i=1}^{n} e^{\left(t^T(a_i\theta_i) + \frac{1}{2}(a_i t^T)\mathbf{\Sigma}_i(a_i t)\right)} \\ &= e^{\left(t^T \sum_{i=1}^{n} a_i\theta_i + \frac{1}{2}\sum_{i=1}^{n} a_i t^T \mathbf{\Sigma}_i t\right)} = e^{\left(t^T \sum_{i=1}^{n} a_i\theta_i + \frac{1}{2} t^T \left(\sum_{i=1}^{n} a_i \mathbf{\Sigma}_i\right) t\right)}. \end{aligned}$$

Dermed har $\sum_{i=1}^{n} a_i Y_i$ følgende fordeling

$$\sum_{i=1}^{n} a_i Y_i \sim N_d\left(\sum_{i=1}^{n} a_i\theta_i,\ \sum_{i=1}^{n} a_i^2 \mathbf{\Sigma}_i\right).$$

■

4.3 Flere sætninger

De følgende sætninger omkring den flerdimensionale normalfordeling finder anvendelse som hjælpesætninger i kapitel 5.

Sætning 4.3.1. *Lad $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$, hvor $X_1, X_2, \cdots, X_m$ er uafhængige og alle har fordelingen $N_d(0, \mathbf{\Sigma})$. Da gælder at $X \sim N_{md}(0, \boldsymbol{I}_m \otimes \mathbf{\Sigma})$*

Bevis

Lad $q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix}$, hvor $q_1, q_2, \cdots, q_m$ er d-dimensionale vektorer. Da gælder ifølge sætning 4.2.1 at $q_i^T X_i \sim N(0, q_i^T \mathbf{\Sigma} q_i)$. Yderligere da $X_1, X_2, \cdots, X_m$ er uafhængige, får man $Cov[X_i, X_j] = \mathbf{O}$ for alle $i \neq j$.

Det følger

$$
\begin{aligned}
&q_i^T X_i \sim N(0, q_i^T \mathbf{\Sigma} q_i) \;\text{ og }\; Cov[X_i, X_j] = \mathbf{O} && \Leftrightarrow \\
&q^T X \sim N\left(0, \sum_{i=1}^{m} q_i^T \mathbf{\Sigma} q_i\right) && \Leftrightarrow \\
&q^T X \sim N(0, q^T(\mathbf{I}_m \otimes \mathbf{\Sigma})q) && \Leftrightarrow \\
&X \sim N_{md}(0, \mathbf{I}_m \otimes \mathbf{\Sigma}).
\end{aligned}
$$

■

Sætning 4.3.2. *Antag at $X_1, X_2, \cdots, X_m$ er uafhængige og alle har fordelingen $N_d(0, \mathbf{\Sigma})$, og lad $\boldsymbol{X} = (X_1 X_2 \cdots X_m)^T$, dvs.*

$$
\boldsymbol{X} = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_m^T \end{pmatrix} = [(X_{ij})] = (X_{1'} X_{2'} \cdots X_{d'})\,.
$$

Da gælder

$$
X_{j'} \sim N_m(0, \sigma_j^2 \boldsymbol{I}_m).
$$

Bevis

Da $X_{ij} \sim N(0, \sigma_j^2)$ gælder, når alle X_i er uafhængige, at $X_{j'} \sim N_m(0, \sigma_j^2 \mathbf{I}_m)$.

■

Sætning 4.3.3. *Hvis c er en m-dimensional vektor og $\boldsymbol{X}$ er defineret som i sætning 4.3.2 gælder*

$$
\boldsymbol{X}^T c \sim N_d(0, ||c||^2 \mathbf{\Sigma})
$$

Bevis

Da $\mathbf{X}^T c$ kan omskrives til $\sum\limits_{i=1}^{m} c_i X_i$ gælder ifølge sætning 4.2.8, at

$$
\mathbf{X}^T c \sim N_d\left(\sum_{i=1}^{n} c_i \theta_i,\; \sum_{i=1}^{n} c_i^2 \mathbf{\Sigma}\right) \sim N_d\left(0\;,\; \sum_{i=1}^{n} c_i^2 \mathbf{\Sigma}\right) = N_d\left(0,\; ||c||^2 \mathbf{\Sigma}\right).
$$

■

Sætning 4.3.4. *Hvis $\{c_1, c_2, \cdots, c_s\}$, $s \leq m$, er ortogale m-dimensionale vektorer, og $\boldsymbol{X}$ er defineret som i sætning 4.3.2, gælder, at de stokastiske variable $\boldsymbol{X}^T c_i$, $i = 1, 2, \cdots, s$, er uafhængige.*

Bevis

Lad $V_i = \mathbf{X}^T c_i, i = 1, 2, \cdots, s$, og lad $c_i = (c_{i1}, c_{i2}, \cdots, c_{im})$, da gælder

$$
\begin{aligned}
Cov[V_i, V_j] &= Cov\left[\sum_{\alpha=1}^{m} c_{i\alpha} X_\alpha,\; \sum_{\beta=1}^{m} c_{j\beta} X_\beta\right] \\
&= \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} c_{i\alpha} c_{j\beta}\; Cov[X_\alpha, X_\beta]
\end{aligned}
$$

Da vektorerne er ortogonale, kan udtrykket omskrives til

$$\begin{aligned}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m} c_{i\alpha}c_{j\beta}\ Cov[X_\alpha, X_\beta] &= \sum_{\alpha=1}^{m} c_{i\alpha}c_{j\alpha}\ Var[X_\alpha] + \mathbf{O}\\ &= c_i^T c_j\ \mathbf{\Sigma}\\ &= \mathbf{O} \qquad \text{for}\ \ i \neq j\end{aligned}$$

Da det ses, at V_i kan skrives som

$$\begin{aligned}V_i &= \sum_{\alpha=1}^{m} c_{i\alpha}X_\alpha = \sum_{\alpha=1}^{m} c_{i\alpha}\mathbf{I}_d X_\alpha\\ &= \begin{pmatrix} c_{i1}\mathbf{I}_d & \dots & c_{im}\mathbf{I}_d\end{pmatrix}\begin{pmatrix} X_1\\ \vdots\\ X_m\end{pmatrix}\\ &= \mathbf{C}X,\end{aligned}$$

og da der ifølge sætning 4.3.1 gælder, at $X \sim N_{md}(0, \mathbf{I}_m \otimes \mathbf{\Sigma})$, får man ud fra sætning 4.2.4, at alle $V_i = \mathbf{X}^T c_i,\ i = 1, 2, \cdots, s$ er uafhængige.

■

Sætning 4.3.5. *Hvis u er en d-dimensional vektor, gælder* $\boldsymbol{X}u \sim N_m(0, \sigma_u^2 \boldsymbol{I}_m)$, *hvor* $\sigma_u^2 = u^T\mathbf{\Sigma}u$.

Bevis

Lad $Y = \mathbf{X}u = \begin{pmatrix} X_1^T u\\ X_2^T u\\ \vdots\\ X_m^T u\end{pmatrix}$.

Da gælder at komponenterne Y_i i Y er uafhængige og alle har fordelingen $N(0, \sigma_u^2)$, hvor

$$\sigma_u^2 = Var[X_i^T u] = Var[u^T X_i] = u^T\mathbf{\Sigma}u.$$

■

Hvis man lader vektoren c være $c = \dfrac{1_m}{m}$, får man $||c||^2 = \dfrac{1}{m}$, og der gælder nu ifølge sætning 4.3.3

$$\overline{X} = \mathbf{X}^T \frac{1_m}{m} \sim N_d\left(0, \frac{\mathbf{\Sigma}}{m}\right).$$

Kapitel 5

Wishartfordelingen

Wishartfordelingen kan til en vis grad siges at være chi-i-anden fordelingen generaliseret til flere dimensioner. Chi-i-anden fordelingen er et specialtilfælde af gammafordelingen $\chi^2_m \sim \Gamma(\frac{m}{2}, 2)$, hvor m er antallet af frihedsgrader. En chi-i-anden fordeling med m frihedsgrader er altså en gammafordeling med formparameter $\frac{m}{2}$ og skalaparameter $\frac{1}{2}$ [Seber, 1984].

5.1 Definitioner

Tæthedsfunktionen for Wishartfordelingen er defineret nedenfor. Der tages udgangspunkt i en positiv definit, symmetrisk og stokastisk matrix $\mathbf{W}$.

Definition 5.1.1. *Lad $\boldsymbol{W} = [(w_{jk})]$ være en matrix med stokastiske variable som elementer. Lad endvidere $\boldsymbol{W}$ være en symmetrisk $d \times d$ matirx, der er positiv definit med sandsynlighed 1, og lad $\boldsymbol{\Sigma}$ være en $d \times d$ matrix, som er positiv definit. Når m er et heltal, sådan at $m \geq d$, siges $\boldsymbol{W}$ at have en Wishartfordeling med m frihedsgrader og parameter $\boldsymbol{\Sigma}$, hvis den simultane tæthedsfunktion af de $\frac{1}{2}d(d+1)$ forskellige elementer i $\boldsymbol{W}$ er*

$$f(w_{11}, w_{12}, \cdots, w_{dd}) = c^{-1}|\boldsymbol{W}|^{\left(\dfrac{m-d-1}{2}\right)} e^{\left(tr\left(\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\right)\right)},$$

hvor

$$c = 2^{\left(\frac{md}{2}\right)}|\boldsymbol{\Sigma}|^{\left(\frac{m}{2}\right)}\Gamma_d\left(\frac{1}{2}m\right),$$

og

$$\Gamma_d\left(\frac{1}{2}m\right) = \pi^{\left(\frac{d(d-1)}{4}\right)} \prod_{i=1}^{d} \Gamma\left(\frac{1}{2}(m+1-i)\right),$$

idet Γ_d er den flerdimensionale gammafunktion.

Hvis $d \times d$ matricen $\mathbf{W}$ er Wishart fordelt, skrives fremover $\mathbf{W} \sim W_d(m, \boldsymbol{\Sigma})$. Det kan i nogle tilfælde være anvendeligt at have en definition af Wishartfordelingen, hvor det ikke er nødvendigt at have kendskab til tæthedsfunktionen. En sådan definition er givet nedenfor.

Definition 5.1.2. *Antag at $X_1, X_2, \cdots, X_m$ er uafhængige og alle har fordelingen $N_d(0, \boldsymbol{\Sigma})$, da har*

$$\boldsymbol{W} = \sum_{i=1}^{m} X_i X_i^T$$

en Wishartfordeling med m frihedsgrader.

Ved at tage afsæt i definition 4.1.2 for den flerdimensionale normalfordeling og ændre restriktionen $|\mathbf{\Sigma}| > 0$ til $|\mathbf{\Sigma}| \geq 0$, kan denne restriktion ligeledes ændres i definition 5.1.2, sådan at $\mathbf{\Sigma}$ her kun kræves at være positiv semidefinit $\mathbf{\Sigma} \geq \mathbf{O}$.
Under antagelse af at $\mathbf{\Sigma}$ er positiv definit og $m \geq d$, kan det vises at $\mathbf{W}$ er positiv definit med sandsynlighed 1. Alternativt vil der ud fra definitionen gælde, at $\mathbf{W}$ er positiv semidefinit, og sandsynligheden for at $|W| = 0$ dermed er større end 0. Selvom det ikke altid er nødvendigt, antages fremadrettet, at $\mathbf{\Sigma}$ er positiv definit. I dette tilfælde kan det vises, at definitionerne 5.1.1 og 5.1.2 er ækvivalente.

5.2 Vigtige egenskaber

Følgende sætning er en analog til sætning 4.2.1 for den flerdimensionale normalfordeling.

Sætning 5.2.1. *Hvis $\boldsymbol{W} \sim W_d(m, \mathbf{\Sigma})$ og $\boldsymbol{C}$ er en $p \times d$ matrix med rang p, gælder at*

$$\boldsymbol{CWC}^T \sim W_p(m, \boldsymbol{CWC}^T).$$

Bevis

Antag at $\mathbf{W}_0 = \sum_{i=1}^{m} X_i X_i^T$, og lad $X_1, X_2, \cdots, X_m$ være uafhængige og alle have fordelingen $N_d(0, \mathbf{\Sigma})$. Dermed gælder at $\mathbf{W}_0 \sim W_d(m, \mathbf{\Sigma})$, sådan at $\mathbf{CW}_0\mathbf{C}^T$ og $\mathbf{CWC}^T$ har samme fordeling. Der gælder desuden, at $\mathbf{CW}_0\mathbf{C}^T$ kan omskrives til

$$\mathbf{CW}_0\mathbf{C}^T = \sum_{i=1}^{m} (\mathbf{C}X_i)(X_i^T\mathbf{C}^T) = \sum_{i=1}^{m} Y_i Y_i^T.$$

Da $Y_i = \mathbf{C}X_i$, gælder ifølge sætning 4.2.1, at $Y_i \sim N_p(0, \mathbf{CWC}^T)$. Dette medfører, jf. definition 5.1.2, at $\mathbf{CW}_0\mathbf{C}^T \sim W_p(m, \mathbf{CWC}^T)$ ■

Sætning 5.2.2. *Hvis u er en d-dimensional vektor, $u \neq 0$, gælder $u^T \boldsymbol{W} u \sim \sigma_u^2 \chi_m^2$, hvor $\sigma_u^2 = u^T \mathbf{\Sigma} u > 0$*

Bevis

Lad $\mathbf{C} = u^T$ og lad $\mathbf{W}_0$ være defineret som i beviset til sætning 5.2.1. Da gælder

$$u^T \mathbf{W}_0 u = \sum_{i=1}^{m} (u^T X_i)^2 = \sum_{i=1}^{m} Y_i^2.$$

Hvor alle Y_i er uafhængige og har fordelingen $N(0, \sigma_u^2)$. Derfor gælder med $u \neq 0$, at

$$\frac{u^T \mathbf{W}_0 u}{\sigma_u^2} \sim \chi_m^2 \Rightarrow u^T \mathbf{W}_0 u \sim \sigma_u^2 \chi_m^2 \ .$$

■

Af sætning 5.2.2 ses, at hvis $u^T = (0, \cdots, 0, 1, 0, \cdots, 0)$, får man $w_{jj} = \sigma_j^2 \chi_m^2$. På trods af dette kaldes Wishartfordelingen ikke for den flerdimensionale chi-i-anden fordeling. Det skyldes, at de maginale fordelinger for $w_{jk}, j \neq k$ ikke er chi-i-anden fordelinger. Ordet flerdimensional anvendes typisk kun i tilfældet, hvor alle endimensionale marginalfordelinger er af samme type.

Sætning 5.2.3. *Lad $\boldsymbol{W} \sim W_d(m, \mathbf{\Sigma})$. Da er middelværdimatricen $E[\boldsymbol{W}] = m\mathbf{\Sigma}$.*

Bevis

Fra definition 5.1.2 har man $\mathbf{W} = \sum_{i=1}^{m} X_i X_i^T = \mathbf{X}^T\mathbf{X}$, hvor $X_i \sim N_d(0, \mathbf{\Sigma}), i = 1, \cdots, m$ er uafhængige. Derfor gælder

$$E[\mathbf{W}] = E[\mathbf{X}^T\mathbf{X}] = E[\mathbf{X}^T\mathbf{I}_m\mathbf{X}].$$

Ifølge sætning 3.2.3 får man

$$\begin{aligned} E[\mathbf{W}] &= tr(\mathbf{I}_m)Var[\mathbf{X}] + E[\mathbf{X}^T]\mathbf{I}_m E[\mathbf{X}] \\ &= m\mathbf{\Sigma} + \mathbf{O}^T\mathbf{I}_m\mathbf{O} = m\mathbf{\Sigma}. \end{aligned}$$

[People-3,2001] ■

Sætning 5.2.4. *Lad $\boldsymbol{W}_1$ og $\boldsymbol{W}_2$ være uafhængige Wishartmatricer med henholdsvis m_1 og m_2 frihedsgrader, og med samme parameter $\mathbf{\Sigma}$. Da gælder at*
$\boldsymbol{W}_1 + \boldsymbol{W}_2 \sim W_d(m_1 + m_2, \;\; \mathbf{\Sigma})$.

Bevis

Da $\mathbf{W}_1$ og $\mathbf{W}_2$ er uafhængige, får man fra definition 5.1.2, at

$$\begin{aligned} \mathbf{W}_1 &= \sum_{i=1}^{m_1} X_i X_i^T, \quad \text{hvor} \quad X_i \sim N_d(0, \mathbf{\Sigma}), \qquad i = 1, \cdots, m_1 \;, \\ \mathbf{W}_2 &= \sum_{i=m_1+1}^{m_1+m_2} X_i X_i^T, \quad \text{hvor} \quad X_i \sim N_d(0, \mathbf{\Sigma}), \quad i = m_1 + 1, \cdots, m_1 + m_2 \;, \end{aligned}$$

hvor alle X_i er uafhængige. Heraf får man

$$\mathbf{W}_1 + \mathbf{W}_2 = \sum_{i=1}^{m_1+m_2} X_i X_i^T, \quad \text{hvor} \quad X_i \sim N_d(0, \mathbf{\Sigma}), \quad i = 1, \cdots, m_1 + m_2 \;,$$

hvor alle X_i er uafhængige. Dermed gælder at $\mathbf{W}_1 + \mathbf{W}_2 \sim W_d(m_1 + m_2, \;\; \mathbf{\Sigma})$.

■

5.3 Generaliseret kvadratisk form

Fra definition 5.1.2 har man $\mathbf{W} = \sum_{i=1}^{m} X_i X_i^T = \mathbf{X}^T\mathbf{X}$.
Ved at opfatte dette som den flerdimensionale version af en kvadratsum, bliver det nærliggende at kigge på det mere generelle udtryk
$\mathbf{X}^T\mathbf{A}\mathbf{X} = \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} X_i X_j^T$, hvor $\mathbf{A} = [(a_{ij})]$ er en symmetrisk $m \times m$ matrix. $\mathbf{X}^T\mathbf{A}\mathbf{X}$ kaldes en generaliseret kvadratisk form. I det endimensionale tilfælde er det matricen $\mathbf{A}$, der afgør egenskaberne for fordelingen. Det samme er gældende for det flerdimensionale tilfælde, faktisk kan $\mathbf{X}^T\mathbf{A}\mathbf{X}$ beskrives ved at kigge på de endimensionale variable $u^T\mathbf{X}^T\mathbf{A}\mathbf{X}u$, se de følgende tre sætninger.

Sætning 5.3.1. *Lad $\boldsymbol{X}^T = (X_1 X_2 \cdots X_m)$, hvor X_i alle har fordelingen $N_d(0, \mathbf{\Sigma})$, og lad $Y = \boldsymbol{X}u$, hvor u er en d-dimensional vektor. Antages desuden, at $\boldsymbol{A}$ er en $m \times m$ matrix med rang r, gælder at $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} \sim W_d(r,\mathbf{\Sigma})$, hvis og kun hvis $Y^T\boldsymbol{A}Y \sim \sigma_u^2\chi_r^2$ for enhvert u, hvor $\sigma_u^2 = u^T\mathbf{\Sigma}u$.*

Bevis

Antag at $\mathbf{W} = \mathbf{X}^T\mathbf{AX} \sim W_d(r,\mathbf{\Sigma})$. Ifølge sætningerne 5.2.1 og 5.2.2, gælder der, at $Y^T\mathbf{A}Y = u^T\mathbf{X}^T\mathbf{AX}u = u^T\mathbf{W}u \sim \sigma_u^2\chi_r^2$.
Antag omvendt, at der for ethvert u gælder $Y^T\mathbf{A}Y \sim \sigma_u^2\chi_r^2$. Ifølge sætning 4.3.5 får man, at $Y \sim N_m(0, \sigma_u^2\mathbf{I}_m)$, og da $\mathbf{A}$ samtidig er en symmetrisk matrix, gælder, at $\mathbf{A}$ er idempotent med rang r, jf. sætning A.0.5. Da $\mathbf{A}$ altså er en projektionsmatrix, kan den, jf. sætning A.0.4, udtrykkes på formen

$$\mathbf{A} = \sum_{i=1}^{r} a_i a_i^T,$$

hvor $a_1, a_2, \cdots, a_r$ er indbyrdes ortonormale egenvektorer hørende til de r egenværdier af $\mathbf{A}$ med værdien 1. Derfor har man

$$\mathbf{X}^T\mathbf{AX} = \sum_{i=1}^{r} \mathbf{X}^T a_i a_i^T \mathbf{X} = \sum_{i=1}^{r} L_i L_i^T,$$

hvor $L_i = \mathbf{X}^T a_i$. Da $||a_i||^2 = 1$, gælder nu ifølge sætningerne 4.3.3 og 4.3.4, at alle L_i er uafhængige og fordelt som $N_d(0,\mathbf{\Sigma})$, og man får nu af definition 5.1.2, at $\mathbf{X}^T\mathbf{AX} \sim W_d(r,\mathbf{\Sigma})$. ■

Sætning 5.3.2. *Lad $\boldsymbol{X}$, $\boldsymbol{A}$ samt Y og u være defineret som i sætning 5.3.1, og lad desuden $\boldsymbol{B}$ være en $m \times m$ matrix med rang s. Da gælder at $\boldsymbol{X}^T\boldsymbol{AX}$ og $\boldsymbol{X}^T\boldsymbol{BX}$ er uafhængige og Wishartfordelte med henholdsvis r og s frihedsgrader, hvis og kun hvis $\frac{Y^T\boldsymbol{A}Y}{\sigma_u^2}$ og $\frac{Y^T\boldsymbol{B}Y}{\sigma_u^2}$ for enhver u er uafhænginge og chi-i-anden fordelte med henholdsvis r og s frihedsgrader.*

Bevis

Antag at $\mathbf{X}^T\mathbf{AX}$ og $\mathbf{X}^T\mathbf{BX}$ er uafhængige og Wishartfordelte. Da gælder, at $\frac{Y^T\mathbf{A}Y}{\sigma_u^2}$ og $\frac{Y^T\mathbf{B}Y}{\sigma_u^2}$ er uafhængige, fordi de er funktioner af uafhængige stokastiske matricer. Det ses desuden af sætning 5.3.1, at $\frac{Y^T\mathbf{A}Y}{\sigma_u^2}$ og $\frac{Y^T\mathbf{B}Y}{\sigma_u^2}$ begge er chi-i-anden fordelte variable.
Antages det omvendt, at $\frac{Y^T\mathbf{A}Y}{\sigma_u^2}$ og $\frac{Y^T\mathbf{B}Y}{\sigma_u^2}$ er uafhænginge og chi-i-anden fordelte med henholdsvis r og s frihedsgrader for en given u, gælder ifølge sætning A.0.5, da $Y \sim N_m(0, \sigma_u^2\mathbf{I}_m)$, at både $\mathbf{A}$ og $\mathbf{B}$ er idempotente med henholdsvis rang r og s. Desuden gælder, jf. sætning A.0.6, at $\mathbf{AB} = \mathbf{O}$. Ifølge sætning A.0.4 kan $\mathbf{A}$ og $\mathbf{B}$ udtrykkes som

$$\mathbf{A} = \sum_{i=1}^{r} a_i a_i^T \quad og \quad \mathbf{B} = \sum_{j=1}^{s} b_j b_j^T,$$

hvor $a_1, a_2, \cdots, a_r$ og $b_1, b_2, \cdots, b_s$ er indbyrdes uafhængige ortonormale egenvektorer hørende til egenværdierne med værdien 1, for henholdsvis $\mathbf{A}$ og $\mathbf{B}$. Da endvidere $\mathbf{A}a_i = a_i$ og $\mathbf{B}b_j = b_j$, får man

$$a_i^T b_j = a_i^T \mathbf{A}^T \mathbf{B} b_j = a_i^T \mathbf{AB} b_j = 0.$$

for alle i og j. Dette betyder, at den forenede mængde af egenvektorer $\{a_1, \cdots, a_r, b_1, \cdots, b_s\}$ er ortonormale. Fra beviset for sætning 5.3.1 følger, at

$$\mathbf{X}^T\mathbf{AX} = \sum_{i=1}^{r} L_i L_i^T \quad \text{og} \quad \mathbf{X}^T\mathbf{BX} = \sum_{j=1}^{s} M_j M_j^T,$$

hvor $L_i = \mathbf{X}^T a_i$ og $M_j = \mathbf{X}^T b_j$. Det følger nu af sætning 4.3.4, at de stokastiske vektorere $L_1, \cdots, L_r, M_1, \cdots, M_s$ alle er indbyrdes uafhængige. Desuden gælder af sætning 4.3.3, da $||a_i||^2 = 1$ og $||b_j||^2 = 1$, at $L_i \sim N_d(0, \mathbf{\Sigma})$ og
$M_j \sim N_d(0, \mathbf{\Sigma})$. Endelig ses af definition 5.1.2, at $\mathbf{X}^T\mathbf{AX}$ og $\mathbf{X}^T\mathbf{BX}$ er uafhængige og Wishartfordelte med henholdsvis r og s frihedsgrader. ∎

Sætning 5.3.3. *Lad $\boldsymbol{X}$ og $\boldsymbol{A}$ samt Y og u være defineret som i de to foregående sætninger. Hvis c er en m-dimensional vektor, gælder at $\boldsymbol{X}^T c$ og $\boldsymbol{X}^T\boldsymbol{AX}$ er uafhængige og fordelt henholdsvis som$N_d(0, ||c||^2\mathbf{\Sigma})$ og $W_d(r,\mathbf{\Sigma})$, hvis og kun hvis $Y^T c$ og $\frac{Y^T\boldsymbol{A}Y}{\sigma_u^2}$ er uafhængige og fordelt henholdsvis som $N(0, ||c||^2\sigma_u^2)$ og χ_r^2 for enhvert u.*

Bevis

Antag at $\mathbf{X}^T c$ og $\mathbf{X}^T\mathbf{AX}$ er uafhængige og fordelt henholdsvis som $N_d(0, ||c||^2\mathbf{\Sigma})$ og $W_d(r,\mathbf{\Sigma})$. Idet $Y = \mathbf{X}u \sim N(0, \sigma_u^2\mathbf{I}_m)$, jf. sætning 4.3.5, gælder der, at
$Y^T c = c^T Y \sim N(0, ||c||^2\sigma_u^2)$, og ifølge sætning 5.2.2 at $Y^T\mathbf{A}Y \sim \sigma_u^2\chi_r^2$. Desuden gælder, at $Y^T c$ og $Y^T\mathbf{A}Y$ er uafhængige, da de er funktioner af uafhængige stokatiske vektorer.
Antag omvendt at $Y^T c$ og $Y^T\mathbf{A}Y$ er uafhængige og fordelt henholdsvis som $N(0, ||c||^2\sigma_u^2)$ og $\sigma_u^2\chi_r^2$ for enhver u. Det ses umiddelbart af sætning 5.3.1, at $\mathbf{X}^T\mathbf{AX} \sim W_d(r,\mathbf{\Sigma})$, og fra sætning 4.3.3 får man $\mathbf{X}^T c \sim N_d(0, ||c||^2\mathbf{\Sigma})$.
Uafhængigheden vises ved at lade $\mathbf{C}$ betegne matricen $\frac{1}{||c||^2}cc^T$. Da gælder, at
$\frac{Y^T\mathbf{C}Y}{c^T c} \sim \sigma_u^2\chi_1^2$. Samtidigt, da både $\mathbf{C}$ og $\mathbf{A}$ er $m \times m$ og symmetriske, gælder ifølge sætning A.0.6, at $\mathbf{AC} = \mathbf{O}$. Desuden kan $\mathbf{A}$ udtrykkes som

$$\mathbf{A} = \sum_{i=1}^{r} a_i a_i^T$$

ifølge sætning A.0.4, hvor $a_1, a_2, \cdots, a_r$ er indbyrdes ortonormale egenvektorer hørende til de r egenværdier af $\mathbf{A}$ med værdien 1. Desuden ses at $\mathbf{C}c = c$, hvoraf der kan udledes

$$\mathbf{AC} = \mathbf{O} \Rightarrow \mathbf{AC}c = \mathbf{O}c \Rightarrow \mathbf{A}c = 0$$

Dette betyder at vektorerne $a_1, \cdots, a_r, c$ alle er ortogonale. Yderligere gælder

$$\mathbf{X}^T\mathbf{AX} = \sum_{i=1}^{r} L_i L_i^T,$$

hvor $L_i = \mathbf{X}^T a_i$ og det ses af sætning 4.3.4, at $\{L_1, \cdots, L_r, \mathbf{X}^T c\}$ alle er uafhængige. Dermed er $\mathbf{X}^T c$ og $\mathbf{X}^T\mathbf{AX}$ uafhængige. ∎

Af de ovenstående tre sætninger 5.3.1, 5.3.2 og 5.3.3 kan følgende udtrykkes:
$\mathbf{X}^T\mathbf{AX} \sim W_d(r, \mathbf{\Sigma})$, hvis og kun hvis $\mathbf{A}^2 = \mathbf{A}$.
De variable $\mathbf{X}^T\mathbf{AX}$ og $\mathbf{X}^T\mathbf{BX}$ er uafhængige, hvis og kun hvis $\mathbf{AB} = \mathbf{O}$.
$\mathbf{X}^T\mathbf{AX}$ og $\mathbf{X}^T c$ er uafhængige og fordelt som $W_d(r, \mathbf{\Sigma})$ og $N_d(0, ||c||^2\mathbf{\Sigma})$, hvis og kun hvis $\mathbf{A}c = 0$ og $\mathbf{A}^2 = \mathbf{A}$.

Kapitel 6

Hotellings T^2 fordeling

Hvis $X \sim N(\mu, \sigma^2)$ og $W \sim \sigma^2 \chi^2_m$ samt X og W er uafhængige gælder

$$T = \frac{\frac{X-\mu}{\sigma}}{\left(\frac{W}{m\sigma^2}\right)^{\frac{1}{2}}} = \frac{X-\mu}{\left(\frac{W}{m}\right)^{\frac{1}{2}}} \sim t_m,$$

hvor t_m er t-fordelingen med m frihedsgrader. Nu ser man, at

$$T^2 = \frac{m\,(X-\mu)^2}{W} = m(X-\mu)W^{-1}(X-\mu) \sim F_{1,m},$$

da sammenhængen mellem t- og F-fordelingen netop er som $t_m^2 \equiv F_{1,m}$ [Seber, 1984].

6.1 Definition

Hotellings T^2 fordeling er en generalisering af overstående.

Definition 6.1.1. *Lad $X \sim N_d(\mu, \mathbf{\Sigma})$ og $\boldsymbol{W} \sim W_d(m, \mathbf{\Sigma})$ være uafhængige. Da siges*

$$T^2 = m(X-\mu)^T \boldsymbol{W}^{-1}(X-\mu)$$

at være T^2-fordelt med parameterne d og m, hvilket noteres som $T^2 \sim T^2_{d,m}$.

6.2 Hjælpesætninger

De følgende fire sætninger introduceres for at kunne bevise sætning 6.3.1, der er angiver sammenhængen mellem Hotellings T^2 fordeling og en F-fordeling.

Sætning 6.2.1. *Antag at $L = (L_1, L_2, \cdots, L_d) \sim N_d(\theta, \mathbf{\Sigma})$, da er den betingede fordeling af L_d, givet $L_1, L_2, \cdots, L_{d-1}$, på formen $N\left(\beta_0 + \sum_{i=1}^{d-1} \beta_i L_i,\ \frac{1}{\sigma^{dd}}\right)$, hvor $[(\sigma^{jk})] = \mathbf{\Sigma}^{-1}$.*

Bevis

Antag, at $L^{(1)} = (L_1, \cdots, L_{d-1})$ og $\theta^{(1)} = (\theta_1, \cdots, \theta_{d-1})$, samt at $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \sigma_{1d} \\ \sigma_{d1}^T & \sigma_{dd} \end{pmatrix}$.

Der gælder ifølge sætning 4.2.7, når $d_2 = 1$, at den betingede fordeling af L_d givet $L^{(1)}$ er normalfordelt med middelværdi

$$\begin{aligned} E[L_d|L^{(1)}] &= \theta_d + \sigma_{d1}^T \mathbf{\Sigma}_{11}^{-1} \left(L^{(1)} - \theta^{(1)}\right) \\ &= \theta_d - \sigma_{d1}^T \mathbf{\Sigma}_{11}^{-1} \theta^{(1)} + \sigma_{d1}^T \mathbf{\Sigma}_{11}^{-1} L^{(1)}, \end{aligned}$$

hvilket kan udtrykkes som

$$E[L_d|L^{(1)}] = \beta_0 + \beta^T L^{(1)}.$$

Da $\mathbf{\Sigma} > \mathbf{O}$ gælder $\mathbf{\Sigma}_{11} > \mathbf{O}$. Man har derfor fra sætning A.0.7, at

$$|\mathbf{\Sigma}| = |\mathbf{\Sigma}_{11}| \left(\sigma_{dd} - \sigma_{d1}^T \Sigma_{11}^{-1} \sigma_{1d}\right)$$

Dermed giver sætning 4.2.7, at

$$\begin{aligned} Var[L_d|L^{(1)}] &= \left(\sigma_{dd} - \sigma_{d1}^T \Sigma_{11}^{-1} \sigma_{1d}\right) \\ &= \frac{|\mathbf{\Sigma}|}{|\mathbf{\Sigma}_{11}|} = \frac{1}{\sigma^{dd}}. \end{aligned}$$

Den sidste omskrivning er mulig, da

$$\sigma^{dd} = (\mathbf{\Sigma}^{-1})_{dd} = \left(\frac{adj(\mathbf{\Sigma})}{|\mathbf{\Sigma}|}\right)_{dd} = \frac{1}{|\mathbf{\Sigma}|}(-1)^{(d+d)}|\mathbf{\Sigma}_{11}| = \frac{|\mathbf{\Sigma}_{11}|}{|\mathbf{\Sigma}|}.$$

■

Sætning 6.2.2. *Betragt den lineære model $Y = \boldsymbol{K}\beta + \mathcal{E}$, hvor $\boldsymbol{K}$ er en $m \times p$ matrix med rang p og $\mathcal{E} \sim N_m(0, \sigma^2 \boldsymbol{I}_m)$.Da gælder, at $Q = Y^T(\boldsymbol{I}_m - \boldsymbol{P})Y \sim \sigma^2 \chi^2_{m-p}$, hvor $\boldsymbol{P} = \boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T$ er en projektionsmatrix, samt at $Q = \dfrac{1}{w^{dd}}$, hvor $\boldsymbol{W} = \begin{pmatrix} \boldsymbol{K}^T \\ Y^T \end{pmatrix} \begin{pmatrix} \boldsymbol{K} & Y \end{pmatrix}$.*

Bevis

Det gælder, at matricen $\mathbf{W}$ kan omskrives til $\mathbf{W} = \begin{pmatrix} \mathbf{K}^T\mathbf{K} & \mathbf{K}^T Y \\ Y^T\mathbf{K} & Y^T Y \end{pmatrix} = \begin{pmatrix} \mathbf{W}_{11} & w_{1d} \\ w_{d1}^T & w_{dd} \end{pmatrix}$, idet det antages, at $\mathbf{W}_{11}$ ikke er stokastisk. Det ses af beviset for sætning 6.2.1, at

$$\frac{1}{w^{dd}} = \frac{|\mathbf{W}|}{|\mathbf{W}_{11}|}$$

Derfor gælder, jf. sætning A.0.7, at

$$\begin{aligned} \frac{1}{w^{dd}} &= \frac{|\mathbf{W}_{11}||w_{dd} - w_{d1}^T \mathbf{W}_{11}^{-1} w_{1d}|}{|\mathbf{W}_{11}|} \\ &= Y^T Y - Y^T\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T Y \\ &= Y^T(\mathbf{I}_m - \mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T)Y \\ &= Y^T(\mathbf{I}_m - \mathbf{P})Y = Q \end{aligned}$$

Da $(\mathbf{I}_m - \mathbf{P})$ er symmetrisk og idempotent, gælder desuden ifølge sætning A.0.5, at $Q = Y^T(\mathbf{I}_m - \mathbf{P})Y \sim \sigma^2\chi^2_{m-p}$. ■

Sætning 6.2.3. *Antag* $\boldsymbol{W} \sim W_d(m, \Sigma)$*, hvor* $m \geq d$*. Da gælder, at* $\dfrac{\sigma^{dd}}{w^{dd}}$ *er fordelt som* χ^2_{m-d+1} *og uafhængig af alle andre elementer* w_{jk} *i* $\boldsymbol{W}$*,* $j, k = 1, 2, \cdots, d-1$.

Bevis
Fra definition 5.1.2 har man $\mathbf{W} = \sum_{i=1}^{m} X_i X_i^T = \mathbf{X}^T \mathbf{X}$, hvor $X_1, X_2, \cdots, X_m$ alle er uafhængige og alle har fordelingen $N_d(0, \mathbf{\Sigma})$. Lad $X_i = (X_{i1}, X_{i2}, \cdots, X_{id})$, da gælder ifølge sætning 6.2.1 at

$$X_{id}|X_i^{(1)} \sim N(0 + \beta^T X_i^{(1)},\ \frac{1}{\sigma^{dd}}), \qquad i = 1, \cdots, m$$
$$\Rightarrow X^{(d)}|\mathbf{K} \sim N_m(\mathbf{K}\beta,\ \frac{1}{\sigma^{dd}}\mathbf{I}_m),$$

hvor $\mathbf{K}^T = \left(X_1^{(1)} \cdots X_m^{(1)}\right)$ og $X^{(d)} = (X_{1d}, X_{2d}, \cdots, X_{md})$

.

Sættes $Y = X^{(d)}$, får man fra sætning 6.2.2, at

$$\frac{1}{w^{dd}}|\mathbf{K} \sim \frac{1}{\sigma^{dd}}\chi^2_{m-(d-1)} = \frac{1}{\sigma^{dd}}\chi^2_{m-d+1}.$$

Fordelingen afhænger altså ikke af $\mathbf{K}$, og derfor gælder, ifølge den flerdimensionale udgave af definition 2.4.2 og sætning 2.5.3, at $\dfrac{1}{w^{dd}}$ har samme fordeling som $\dfrac{1}{w^{dd}}|\mathbf{K}$, dvs.

$$\frac{1}{w^{dd}} \sim \frac{1}{\sigma^{dd}}\chi^2_{m-d+1} \Rightarrow \frac{\sigma^{dd}}{w^{dd}} \sim \chi^2_{m-d+1}.$$

Dermed ses også, at $\dfrac{\sigma^{dd}}{w^{dd}}$ er uafhængig af $\mathbf{W}_{11} = \mathbf{K}^T\mathbf{K}$. ■

Sætning 6.2.4. *Antag* $\boldsymbol{W} \sim W_d(m, \Sigma)$ *hvor* $m \geq d$*. Da gælder for enhvert* $u \neq 0$*, at* $\dfrac{u^T \mathbf{\Sigma}^{-1} u}{u^T \boldsymbol{W}^{-1} u}$ *er fordelt som* χ^2_{m-d+1}.

Bevis
Lad $\mathbf{U}$ være en $d \times d$ ortogonalmatrix, hvor den sidste række er $\dfrac{1}{||u||}u^T$. Det ses at $\mathbf{UWU}^T \sim W_d(m, \mathbf{U\Sigma U}^T)$ ifølge sætning 5.2.1. Når $\mathbf{U}$ er en ortogonalmatrix gælder $\mathbf{U}^T = \mathbf{U}^{-1}$, og dermed også at $(\mathbf{U\Sigma U}^T)^{-1} = \mathbf{U\Sigma}^{-1}\mathbf{U}^T$, samt $(\mathbf{UWU}^T)^{-1} = \mathbf{UW}^{-1}\mathbf{U}^T$. Man får derfor

$$\frac{u^T \mathbf{\Sigma}^{-1} u}{u^T \mathbf{W}^{-1} u} = \frac{\left(\frac{1}{||u||}u^T\right)\mathbf{\Sigma}^{-1}\left(\frac{1}{||u||}u\right)}{\left(\frac{1}{||u||}u^T\right)\mathbf{W}^{-1}\left(\frac{1}{||u||}u\right)} = \frac{(\mathbf{U\Sigma}^{-1}\mathbf{U}^T)_{dd}}{(\mathbf{UW}^{-1}\mathbf{U}^T)_{dd}} =$$
$$\frac{((\mathbf{U\Sigma U}^T)^{-1})_{dd}}{((\mathbf{UWU}^T)^{-1})_{dd}} = \frac{(\mathbf{U\Sigma U}^T)^{dd}}{(\mathbf{UWU}^T)^{dd}}.$$

Der gælder, jf. sætning 6.2.3, at $\dfrac{(\mathbf{U\Sigma U}^T)^{dd}}{(\mathbf{UWU}^T)^{dd}} \sim \chi^2_{m-d+1}$ og dermed også at $\dfrac{u^T \mathbf{\Sigma}^{-1} u}{u^T \mathbf{W}^{-1} u} \sim \chi^2_{m-d+1}$. ■

6.3 Egenskaber

Hotellings T^2-fordelingen afhænger af to parametre, d og m, og noteres $T^2 \sim T^2_{d,m}$.

Sætning 6.3.1. *Antag, at* $T^2 = mY^T\boldsymbol{W}^{-1}Y \sim T^2_{d,m}$*, hvor* $Y \sim N_d(0, \boldsymbol{\Sigma})$*, og at* $\boldsymbol{W} \sim W_d(m, \Sigma)$*, samt at* Y *og* $\boldsymbol{W}$ *er uafhængige. Da gælder*

$$\frac{m-d+1}{d}\frac{T^2}{m} \sim F_{d,\ m-d+1}.$$

Bevis
Lad H betegne $\dfrac{Y^T\boldsymbol{\Sigma}^{-1}Y}{Y^T\mathbf{W}^{-1}Y}$ og G betegne $Y^T\boldsymbol{\Sigma}^{-1}Y$, da får man ved omskrivning

$$\frac{T^2}{m} = Y^T\mathbf{W}^{-1}Y = \frac{Y^T\boldsymbol{\Sigma}^{-1}Y}{\frac{Y^T\boldsymbol{\Sigma}^{-1}Y}{Y^T\mathbf{W}^{-1}Y}} = \frac{G}{H}$$

Fra sætning 6.2.4 får man $H|Y \sim \chi^2_{m-d+1}$. Man ser, at den betingede fordeling ikke afhænger af Y. Dermed har H samme fordeling som $H|Y$, dvs. $H \sim \chi^2_{m-d+1}$, og H og Y er uafhængige. Med $\theta = 0$ gælder, jf. sætning 4.2.5, at $G \sim \chi^2_d$. Det følger desuden, at G er uafhængig af H. Derfor er $\dfrac{T^2}{m}$ forholdet mellem to chi-i-anden fordelte variable, hvoraf følger

$$\frac{\frac{G}{d}}{\frac{H}{m-d+1}} = \frac{m-d+1}{d}\frac{T^2}{m} \sim F_{d,\ m-d+1}.$$

■

Sætning 6.3.2. *Hvis* $X \sim N_d(\mu, \lambda^{-1}\boldsymbol{\Sigma})$ *og* $\boldsymbol{W} \sim W_d(m, \Sigma)$ *er uafhængige, gælder*

$$T^2 = \lambda m(X-\mu)^T\boldsymbol{W}^{-1}(X-\mu) \sim T^2_{d,m}$$

Bevis
Lad Y være $Y = \sqrt{\lambda}(X-\mu) \sim N_d(0, \Sigma)$, da får man ved omskrivning

$$T^2 = \lambda m(X-\mu)^T\mathbf{W}^{-1}(X-\mu) = mY^T\mathbf{W}^{-1}Y.$$

Det følger derfor af sætning 6.3.1, at $T^2 \sim \dfrac{md}{m-d+1}F_{d,\ m-d+1} \equiv T^2_{d,m}$. ■

Kapitel 7

Den flerdimensionale betafordeling

Hvis X og Y stokastiske variable, der begge er chi-i-anden fordelte med henholdsvis n og m frihedsgrader, gælder at $\frac{X}{Y}$ og $\frac{X}{X+Y}$ er henholdvis F- og betafordelte. I det flerdimensionale tilfælde får Wishartfordelingen samme betydning som chi-i-anden fordelingen for det endimensionale. Der findes dog flere genereliseringer, der fører til flerdimensionale analoger af F- og betafordelingerne.
Med afsæt i den almindelige betafordeling, vil dette kapitel beskrive en flerdimensionale betafordeling [Seber, 1984].

7.1 Udledning og definition

Lad $H \sim \sigma^2\chi_{m_H}$ og $E \sim \sigma^2\chi_{m_E}$ være uafhængige. Da er tæthedsfunktionerne for $T = \frac{H}{E}$ og $V = T(1+T) = \frac{H}{E+H}$ henholdsvis

$$f(t) = \frac{1}{B(m_H/2,\ m_E/2)} \frac{t^{(m_H/2-1)}}{(1+t)^{(m_E+m_H)/2}}, \qquad 0 \leq t < \infty\ ,$$

og

$$g(v) = \frac{1}{B(m_H/2,\ m_E/2)} v^{m_H/2-1}(1-v)^{m_E/2-1}, \qquad 0 \leq v \leq 1\ ,$$

hvor $B(a,b)$ er betafunktionen

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Det er hensigtsmæssigt at bruge notationen $V \sim B_{m_H/2,\ m_E/2}$, og V siges at have en type 1 betafordeling med henholdsvis $\frac{1}{2}m_H$ og $\frac{1}{2}m_H$ frihedsgrader. Det bemærkes, at $\dfrac{m_E T}{m_H} \sim F_{m_h, m_e}$, og T siges at have en type 2 betafordeling med henholdsvis $\frac{1}{2}m_H$ og $\frac{1}{2}m_H$ frihedsgrader. Desuden gælder

$$1 - B_{a,b} \sim B_{b,a} \sim \left(1 + \frac{2a}{2b}F_{2a,2b}\right)^{-1}$$

De ovenstående resultater kan genereliseres til tilfældet, hvor $\mathbf{H}$ og $\mathbf{E}$ er matricer med uafhængige Wishartfordelnger, altså $\mathbf{H} \sim W_d(m_H, \mathbf{\Sigma})$ og $\mathbf{E} \sim W_d(m_E, \mathbf{\Sigma})$, hvor

$m_H, m_E \geq d$.
Da $\mathbf{H}$, $\mathbf{E}$ og dermed også $\mathbf{H}+\mathbf{E}$ er positive definite matricer med sandsynlighed 1, kan der ifølge sætning A.0.1 findes to positive definite matricer $\mathbf{T}$ og $\mathbf{V}$

$$\mathbf{T} = \mathbf{E}^{-\frac{1}{2}}\mathbf{H}\mathbf{E}^{-\frac{1}{2}}.$$
$$\mathbf{V} = (\mathbf{E}+\mathbf{H})^{-\frac{1}{2}}\mathbf{H}(\mathbf{E}+\mathbf{H})^{-\frac{1}{2}},$$

hvor både $\mathbf{E}$ og $\mathbf{H}$ er symmetriske.

Sætning 7.1.1. *Den simultane tæthedsfunktion af de $\frac{1}{2}d(d+1)$ forskellige elementer i $\boldsymbol{V}$, altså $g(\boldsymbol{V}) = g(v_{11}, v_{12}, \cdots, v_{dd})$ er givet som*

$$g(\boldsymbol{V}) = \frac{1}{B_d(m_H/2,\ m_E/2)}|\boldsymbol{V}|^{(m_H-d-1)/2}|\boldsymbol{I}_d - \boldsymbol{V}|^{(m_E-d-1)/2}, \qquad \boldsymbol{O} < \boldsymbol{V} < \boldsymbol{I}_d\ ,$$

hvor $B_d(a,b)$ er den flerdimensionale betafunktion

$$B_d(a,b) = \frac{\Gamma_d(a)\Gamma_d(b)}{\Gamma_d(a+b)}, \qquad d \leq 2a, 2b$$

og $\Gamma_d(a)$ er den flerdimensionale gammafunktion, som anført i definition 5.1.1. Det antages desuden at $d \leq m_H, m_E$.

Bevis
Da $\mathbf{H}$ og $\mathbf{E}$ er uafhængige, gælder at deres simultane tæthedsfunktion er produktet af de marginale tæthedsfunktioner. Derfor fås, jf. definition 5.1.1, at

$$\begin{aligned} f(\mathbf{H},\mathbf{E}) &= \frac{1}{c_H}|\mathbf{H}|^{\frac{m_H-d-1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}\mathbf{H}\right)} \frac{1}{c_E}|\mathbf{E}|^{\frac{m_E-d-1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}\mathbf{E}\right)} \\ &= \frac{1}{c_H c_E}|\mathbf{H}|^{\frac{m_H-d-1}{2}}|\mathbf{E}|^{\frac{m_E-d-1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}(\mathbf{H}+\mathbf{E})\right)}. \end{aligned}$$

Herefter foretages et variabelskift ved at definere $\mathbf{V} = (\mathbf{H}+\mathbf{E})^{-\frac{1}{2}}\mathbf{H}(\mathbf{H}+\mathbf{E})^{-\frac{1}{2}}$ og $\mathbf{Z} = \mathbf{H}+\mathbf{E}$. Det følger af sætning A.0.9, at Jacobideterminanten ved dette variabelskifte er $\dfrac{\partial(\mathbf{H},\mathbf{E})}{\partial(\mathbf{V},\mathbf{Z})} = |\mathbf{Z}|^{\frac{d+1}{2}}$. Dermed kan den simultane tæthedsfunktion for de øvre trekantelementer i $\mathbf{V}$ og $\mathbf{Z}$ findes som

$$\begin{aligned} h(\mathbf{V},\mathbf{Z}) &= \frac{1}{c_H c_E}|\mathbf{Z}^{\frac{1}{2}}\mathbf{V}\mathbf{Z}^{\frac{1}{2}}|^{\frac{m_H-d-1}{2}}|\mathbf{Z} - \mathbf{Z}^{\frac{1}{2}}\mathbf{V}\mathbf{Z}^{\frac{1}{2}}|^{\frac{m_E-d-1}{2}}|\mathbf{Z}|^{\frac{d+1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}\mathbf{Z}\right)} \\ &= \frac{1}{c_H c_E}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d - \mathbf{V}|^{\frac{m_E-d-1}{2}}|\mathbf{Z}|^{\frac{m_H+m_E-d-1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}\mathbf{Z}\right)}. \end{aligned}$$

Det ses, at den simultane tæthedsfunktion af de $\frac{1}{2}d(d+1)$ forskellige elementer i $\mathbf{V}$ kan findes som

$$\begin{aligned} g(\mathbf{V}) &= \int_{\mathbb{R}^{\frac{d(d+1)}{2}}} h(\mathbf{V},\mathbf{Z}) \ d\Omega_z \\ &= \frac{c_{H+E}}{c_H c_E}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d - \mathbf{V}|^{\frac{m_E-d-1}{2}}. \\ &\qquad \int_{\mathbb{R}^{\frac{d(d+1)}{2}}} \frac{1}{c_{H+E}}|\mathbf{Z}|^{\frac{m_H+m_E-d-1}{2}} e^{tr\left(-\frac{1}{2}\Sigma^{-1}\mathbf{Z}\right)} \ d\Omega_z. \end{aligned}$$

Da udtrykket under integralet en tæthedsfunktion for en Wishartfordeling, er integralets værdi 1. Det gælder derfor, at

$$g(\mathbf{V}) = \frac{c_{H+E}}{c_H c_E}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d - \mathbf{V}|^{\frac{m_E-d-1}{2}}.$$

Med inspiration fra definition 5.1.1 kan udtrykket skrives som

$$\begin{aligned}
g(\mathbf{V}) &= \frac{2^{\frac{(m_H+m_E)d}{2}}|\boldsymbol{\Sigma}|^{\frac{m_H+m_E}{2}}\Gamma_d\left(\frac{m_H+m_E}{2}\right)}{2^{\frac{m_Hd}{2}}|\boldsymbol{\Sigma}|^{\frac{m_H}{2}}\Gamma_d\left(\frac{m_H}{2}\right)2^{\frac{m_Ed}{2}}|\boldsymbol{\Sigma}|^{\frac{m_E}{2}}\Gamma_d\left(\frac{m_E}{2}\right)}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d-\mathbf{V}|^{\frac{m_E-d-1}{2}} \\
&= \frac{\Gamma_d\left(\frac{m_H+m_E}{2}\right)}{\Gamma_d\left(\frac{m_H}{2}\right)\Gamma_d\left(\frac{m_E}{2}\right)}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d-\mathbf{V}|^{\frac{m_E-d-1}{2}} \\
&= \frac{1}{B_d\left(\frac{m_H}{2},\frac{m_E}{2}\right)}|\mathbf{V}|^{\frac{m_H-d-1}{2}}|\mathbf{I}_d-\mathbf{V}|^{\frac{m_E-d-1}{2}}.
\end{aligned}$$

Afslutningsvis skal det vises, at $\mathbf{O} < \mathbf{V} < \mathbf{I}_d$ med sandsynlighed 1. Det gælder jf. sætning A.0.1 at $\mathbf{V} > \mathbf{O}$, desuden ses at

$$\begin{aligned}
\mathbf{I}_d - \mathbf{V} &= (\mathbf{E}+\mathbf{H})^{-\frac{1}{2}}(\mathbf{E}+\mathbf{H}-\mathbf{H})(\mathbf{E}+\mathbf{H})^{-\frac{1}{2}} \\
&= (\mathbf{E}+\mathbf{H})^{-\frac{1}{2}}\mathbf{E}(\mathbf{E}+\mathbf{H})^{-\frac{1}{2}} > \mathbf{O}.
\end{aligned}$$

Heraf følger altså $\mathbf{O} < \mathbf{V} < \mathbf{I}_d$.

■

Tæthedsfunktionen for $\mathbf{V}$ afhænger ikke af $\boldsymbol{\Sigma}$. I modsætning hertil afhænger tæthedsfunktionen for $\mathbf{T}$ af $\boldsymbol{\Sigma}$. For $\boldsymbol{\Sigma} = \mathbf{I}_d$ kan det vises, at

$$f(\mathbf{T}) = \frac{1}{B_d(m_H/2,\ m_E/2)}|\mathbf{T}|^{(m_H-d-1)/2}|\mathbf{I}_d-\mathbf{T}|^{-(m_E+m_H)/2}, \qquad \mathbf{T} > \mathbf{O}.$$

Som i de endimensionale tilfælde siges denne funktion at være tæthedsfunktionen for en flerdimensional type II betafordeling, mens $\mathbf{V}$ siges at have en d-dimensional type I betafordeling med henholdsvis $\frac{1}{2}m_H$ og $\frac{1}{2}m_E$ frihedsgrader.

Kapitel 8

Afrunding

I begyndelsen af projektet blev simultane fordelinger beskrevet med udgangspunkt i todimensionale stokastiske variable. Det er bl.a. blevet uddybet, hvad der gælder for diskrete, kontinuerte samt betingede fordelinger. Videre er det blevet beskrevet, hvad der gælder for uafhængige stokastiske variable, middelværdi og varians samt kovarians og korrelation. Med eksempler er det desuden vist, hvordan de forskellige fordelinger og operatorer kan anvendes.
Disse begreber er forsøgt udvidet og således videreført til flerdimensionale fordelinger. Fokusset her har været at give en detaljeret gennemgang af matematikken bag den flerdimensionale normalfordeling og Wishartfordelingen samt kort at belyse Hotellings T^2 fordeling og den flerdimensionale betafordeling. Flere af disse fordelinger er analoge til endimensionale fordelinger.

Litteraturliste

[Econ,2004]
http://www2.econ.iastate.edu/classes/econ671/hallam/documents/mpd.pdf;
Sidst besøgt 29. marts 2015.

[Gut, 2009]
Gut, A.;2009; An Intermediate Course in Probability;
Springer Science + Business Media, LCC; ISBN: 978-1-4419-0161-3.

[Mardia et al., 1979]
Mardia, K. V., Kent, J. T. og Bibby, J. M.;1979; Multivariate Analysis;
Academic Press; ISBN: 9780124712522.

[Olofsson & Andersson, 2012]
Olofsson, P. & Anderson, M.; 2012; Probability, Statistics, and Stochastic Processes;
2. udgave; John Wiley & Sons, Inc; ISBN: 978-0-470-88974-9.

[People-1,2001]
http://people.math.aau.dk/ br/mat4(nu6)-stat-2001/note2.pdf; 2001;
The Multidimensional Normal Distribution; Rosbjerg, B.; Sidst besøgt 29. marts 2015.

[People-2,2001]
http://people.math.aau.dk/ br/mat4(nu6)-stat-2001/loesn2.pdf; 2001;
Rosbjerg, B.; Sidst besøgt 29. marts 2015.

[People-3,2001]
http://people.math.aau.dk/ br/mat4(nu6)-stat-2001/loesn4.pdf; 2001;
Rosbjerg, B.; Sidst besøgt 29. marts 2015.

[Seber, 1984]
Seber, G.A.F.;1984; Multivariate Observations; John Wiley & Sons, Inc;
ISBN: 0-471-88104-X.

Notationsforklaring

Vektorer noteres med små bogstaver, således indeholder en d-dimensional vektor a komponeterne $a_1, a_2, \cdots, a_d$.
Stokastiske variable noteres med store bogstaver, og værdierne af disse noteres med det tilhørende lille bogstav, eksempelvis Y og y.
Stokastiske vektorer betegnes ligeledes med store bogstaver. En d-dimensional stokastisk vektor X har altså de stoakstiske variable $X_1, X_2, \cdots, X_d$ som komponenter, og værdien af X noteres $x = (x_1, x_2, \cdots, x_d)$.
Matricer noteres med store bogstaver og fed skrift. Hvis en $m \times n$ matrix $\mathbf{A}$ har elementerne a_{ij} skrives $\mathbf{A} = [a_{ij}]$.
En stokastisk matrix noteres $\mathbf{X} = [X_{ij}]$, dvs.

$$\mathbf{X} = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1d} \\ X_{21} & X_{22} & \cdots & X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nd} \end{pmatrix}.$$

Hver række i matricen er altså en d-dimensional stokastisk vektor, og alle elementerne i matricen er stokastiske variable. Eksempelvis er X_{ij} det j'te komponent i den i'te stokastiske vektor. Søjlevektorerne i $\mathbf{X}$ betegnes som $X_{j'}$, hvor $j = 1, \cdots, d$.
Det er vigtigt, at læseren skelner mellem notationen $Z = (Z_1, Z_2, \cdots, Z_d)$, der er en d-dimensional stokastisk vektor, og $\mathbf{Z} = (Z_{1'}\ Z_{2'}\ \cdots\ Z_{d'})$, som er en $n \times d$ stokastisk matrix. Det bemærkes, at Z_j betegner det j'te komponent i en stokastisk vektor Z, og $Z_{j'}$ betegner den j'te søjlevektor i den stokastiske matrix $\mathbf{Z}$.

Bilag A

Matrixalgebra

Dette bilag indeholder vigtige sætninger om matrixalgebra, som anvendes i kaptilerne 4 og 5.

Sætning A.0.1. *Lad $\boldsymbol{A}$ være en positiv definit matrix og lad $\boldsymbol{C}$ være en $p \times n$ matix med rang p. Da gælder, at $\boldsymbol{CAC}^T$ er positiv definit*

Bevis
Da $\mathbf{A}$ er positiv definit, gælder $y^T\mathbf{A}y > 0$ for alle $y \neq 0$. Hvis man lader vektoren y være $y = \mathbf{C}^T x$, får man

$$y^T\mathbf{A}y = x^T\mathbf{CAC}^T x = 0 \Rightarrow y = \mathbf{C}^T x = 0.$$

Da alle søjlerne i $\mathbf{C}^T$ er lineært uafhængige, gælder at $x = 0$. Heraf ses, at $x^T\mathbf{CAC}^T x > 0$ for alle $x \neq 0$. Dermed er matricen $\mathbf{CAC}^T$ positiv definit.

■

Sætning A.0.2. *Enhver symmetrisk $n \times n$ matrix $\boldsymbol{A}$ kan skrives som*

$$\boldsymbol{A} = \boldsymbol{TDT}^T = \sum_{i=1}^{n} \lambda_i t_i t_i^T,$$

hvor $\boldsymbol{D}$ er en diagonalmatrix indeholdende egenværdierne for $\boldsymbol{A}$, og $\boldsymbol{T}$ er en ortogonal matrix, hvis søjler er ortonormale egenvektorer.

Bevis
Antag, at der findes ortonormale vektorer, så

$$t_i^T\mathbf{A}t_j = \lambda_i t_i^T t_j = \begin{cases} \lambda_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases},$$

hvilket kan udtrykkes på matrixform som

$$\mathbf{T}^T\mathbf{AT} = \mathbf{D}.$$

Hvis udtrykket multipliceres fra venstre med $\mathbf{T}$ og fra højre med $\mathbf{T}^T$, fås

$$\mathbf{TT}^T\mathbf{ATT}^T = \mathbf{TDT}^T \Rightarrow \mathbf{A} = \mathbf{TDT}^T.$$

Da similære matricer har samme egenværdier, gælder at $\mathbf{A}$ og $\mathbf{D}$ har samme egenværdier, hvilket betyder, at elementerne i $\mathbf{D}$ er egenværdierne for $\mathbf{A}$.

Der skal altså findes en ortonormal basis af egenvektorer. Det bemærkes, at hvis $\lambda_i \neq \lambda_j$ er forskellige egenværdier med tilhørende egenvektorer y og x, gælder $\lambda_i y^T x = \lambda_i x^T y = x^T \mathbf{A} y = y^T \mathbf{A} x = \lambda_j y^T x$, hvoraf det ses, at $y^T x = 0$. Dermed gælder, at egenvektorerne tilhørende forskellige egenværdier for en symmetrisk matrix er ortogonale.
Antag at $\mathbf{A}$ har k forskellige egenværdier med egenrummene $\mathbf{E}_1, \mathbf{E}_2, \cdots, \mathbf{E}_k$, der har dimensionerne $r_1, r_2, \cdots, r_k$. Det følger af ovenstående, at egenrummene $\mathbf{E}_1, \mathbf{E}_2, \cdots, \mathbf{E}_k$ er ortogonale. Lad $r = \sum_{j=1}^{k} r_j$, da findes ortonormale vektorer med indeks

$$\sum_{i=1}^{j-1} r_i + 1, \sum_{i=1}^{j-1} r_i + 2, \cdots, \sum_{i=1}^{j} r_i,$$

som danner en basis for $\mathbf{E}_j$. Der gælder desuden, at r_j er mindre end eller lig med den algebraiske multiplicitet af den tilhørende egenværdi. Derfor kan det antages at

$$\mathbf{A} e_i = \lambda_i e_i, \quad i = 1, 2, \cdots, r,$$

hvor $r \leq n$. Hvis $r = n$, og man sætter $t_i = e_i$ får man $\mathbf{A} t_i = \lambda_i t_i$, hvilket beviser sætningen.
Hvis derimod $r < n$, vil vi vise, at dette fører til en modstrid. Det antages, at $\lambda_i > 0$ for alle i. Hvis ikke, sættes $\mathbf{A} := \mathbf{A} + c\mathbf{I}, \ \ c > 0$, som har samme egenvektorer.

Sæt

$$\mathbf{B} = \mathbf{A} - \sum_{i=1}^{r} \lambda_i e_i e_i^T,$$

hvilket giver

$$tr\mathbf{B} = tr\mathbf{A} - \sum_{i=1}^{r} \lambda_i e_i^T e_i = \sum_{i=r+1}^{n} \lambda_i > 0$$

Dermed har $\mathbf{B}$ mindst en positiv egenværdi α. Antag, at x er den tilhørende egenvektor, da gælder

$$\begin{aligned} e_j^T \alpha x = e_j^T \mathbf{B} x &= e_j^T \left(\mathbf{A} - \sum_{i=1}^{r} \lambda_i e_i e_i^T \right) x \\ &= \left(\lambda_j e_j^T - \sum_{i=1}^{r} \lambda_i e_j^T e_i e_i^T \right) x \\ &= \lambda_j e_j^T x - \lambda_j e_j^T x = 0, \quad \text{for} \ \ 1 \leq j \leq r \end{aligned}$$

Ovenstående viser at x er ortogonal på e_j, $j = 1, 2, \cdots, r$. Hvoraf følger, at

$$\begin{aligned} \alpha x = \mathbf{B} x &= \left(\mathbf{A} - \sum_{i=1}^{r} \lambda_i e_i e_i^T \right) x \\ &= \mathbf{A} x - \sum_{i=1}^{r} \lambda_i e_i e_i^T x \\ &= \mathbf{A} x - \sum_{i=1}^{r} \lambda_i (e_i^T x) e_i = \mathbf{A} x \end{aligned}$$

Det ses at x er en egenvektor for $\mathbf{A}$, dvs. der findes et i så $\alpha = \lambda_i$. Dermed er x en linearkombination af e_i-erne, men det er i modstrid med, at x er ortogonal på e_i-erne.

[Mardia et al., 1979] ■

Sætning A.0.3. *Hvis $\boldsymbol{P}$ er en $n \times n$ symmetrisk matrix, gælder der, at $\boldsymbol{P}$ er en idempotent matrix med rang r, hvis og kun hvis den har r egenværdier, som er 1, og $n - r$ egneværdier, der er 0.*

Bevis

Antag, at $\mathbf{P}$ er idempotent med rang r, og lad $\mathbf{P}x = \lambda x$, hvor $x \neq 0$. Dette medfører, at $x^T\mathbf{P}x = x^T\lambda x$. Da λ er en skalar, får man

$$x^T\lambda x = \lambda x^T x = x^T\mathbf{P}x = x^T\mathbf{P}^2 x = (\mathbf{P}x)^T(\mathbf{P}x) = \lambda x^T\lambda x = \lambda^2 x^T x$$

Det ses at $\lambda x^T x = \lambda^2 x^T x$, hvilket videre giver

$$0 = \lambda^2 x^T x - \lambda x^T x = \lambda(\lambda x^T x - x^T x) = \lambda(\lambda - 1)x^T x$$

Da $x^T x \neq 0$, gælder $\lambda(\lambda - 1) = 0$, hvoraf det ses at egenværdierne er 1 eller 0. Yderligere, da $\mathbf{P}$ har rang r, og da antallet af egenværdier forskellig fra 0 i en symmetrisk matrix er lig med rangen, gælder at $\mathbf{P}$ har r egenværdier, der er 1, og $n-r$ egenværdier, som er 0.

Omvendt antages, at egenværdierne til $\mathbf{P}$ er 1 eller 0. Det følger af sætning A.0.2, da $\mathbf{P}$ er en symmetrisk $n \times n$ matrix, at der eksisterer en ortogonalmatrix $\mathbf{T}$ sådan at

$$\begin{aligned}\mathbf{T}^T\mathbf{P}\mathbf{T} &= diag(\lambda_1, \lambda_2, \cdots, \lambda_n) \\ &= \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix} \\ &= \mathbf{\Lambda}\end{aligned}$$

Dette viser, at $\mathbf{T}\mathbf{T}^T\mathbf{P}\mathbf{T}\mathbf{T}^T = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^T$ og dermed $\mathbf{P} = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^T$, hvoraf det følger

$$\mathbf{P}^2 = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^T\mathbf{T}\mathbf{\Lambda}\mathbf{T}^T = \mathbf{T}\mathbf{\Lambda}^2\mathbf{T}^T = \mathbf{T}\mathbf{\Lambda}\mathbf{T}^T = \mathbf{P}.$$

Altså er $\mathbf{P}$ idempotent. Rangen af $\mathbf{P}$, der også er rangen af $\mathbf{T}\mathbf{\Lambda}\mathbf{T}^T$, kan findes som

$$rang(\mathbf{P}) = rang(\mathbf{\Lambda}) = r,$$

da $\mathbf{T}$ ikke er singulær. ■

Iøvrigt gælder følgende sætninger om idempotente matricer.

Sætning A.0.4. *Hvis $\boldsymbol{P}$ er en projektionsmatrix med rang r, kan den udtrykkes på formen*

$$\boldsymbol{P} = \sum_{i=1}^{r} t_i t_i^T$$

hvor $t_1, t_2, \cdots, t_r$ er indbyrdes ortonormale.

Bevis
Antag at $\mathbf{P}$ er en $n \times n$ projektionsmatrix. Det ses af sætning A.0.2, at der findes en diagonalmatrix $\mathbf{T} = (t_1 \ t_2 \ \cdots \ t_n)$, sådan at

$$\mathbf{A} = \mathbf{TDT}^T = \sum_{i=1}^{n} \lambda_i t_i t_i^T,$$

hvor λ_i er egenværdierne af $\mathbf{P}$. Fordi $\mathbf{P}$ har rang r, følger af sætning A.0.3, at $\mathbf{P}$ har r egenværdier, der er 1, og at de resterende egenværdier har værdien 0. Heraf får man

$$\mathbf{P} = \sum_{i=1}^{r} t_i t_i^T.$$

■

Sætning A.0.5. *Lad $X \sim N_d(0, \sigma^2 \boldsymbol{I}_d)$, og $\boldsymbol{P}$ være en symmetrisk $d \times d$ matix. Da gælder $X^T \boldsymbol{P} X \sim \sigma^2 \chi_r^2$, hvis og kun hvis $\boldsymbol{P}$ er idempotent med rang r.*

Bevis

Først antages at $X^T \mathbf{P} X \sim \sigma^2 \chi_r^2$. Den momentfrembringende funktion kan findes til

$$\begin{aligned}
M_{X^T\mathbf{P}X}(t) &= E[e^{(tX^T\mathbf{P}X)}] \\
&= \int_{\mathbb{R}_d} e^{tX^T\mathbf{P}X}(2\pi)^{-\frac{d}{2}}|\sigma^2\mathbf{I}_d|^{-\frac{1}{2}}e^{\left(-\frac{1}{2}X^T\sigma^2\mathbf{I}_dX\right)}d\Omega \\
&= \int_{\mathbb{R}_d} (2\pi)^{-\frac{d}{2}}|\sigma^2\mathbf{I}_d|^{-\frac{1}{2}}e^{\left(-\frac{1}{2}X^T((\sigma^2\mathbf{I}_d)^{-1}-2t\mathbf{P})X\right)}d\Omega \\
&= |\mathbf{I}_d - 2t\mathbf{P}\sigma^2\mathbf{I}_d|^{-\frac{1}{2}} \int_{\mathbb{R}_d} (2\pi)^{-\frac{d}{2}}|\sigma^2\mathbf{I}_d|^{-\frac{1}{2}}|\mathbf{I}_d - 2t\mathbf{P}\sigma^2\mathbf{I}_d|^{\frac{1}{2}}e^{\left(-\frac{1}{2}X^T((\sigma^2\mathbf{I}_d)^{-1}-2t\mathbf{P})X\right)}d\Omega \\
&= |\mathbf{I}_d - 2t\mathbf{P}\sigma^2\mathbf{I}_d|^{-\frac{1}{2}} \int_{\mathbb{R}_d} (2\pi)^{-\frac{d}{2}}(|(\sigma^2\mathbf{I}_d)^{-1} - 2t\mathbf{P}|^{-1})^{-\frac{1}{2}}e^{\left(-\frac{1}{2}X^T((\sigma^2\mathbf{I}_d)^{-1}-2t\mathbf{P})X\right)}d\Omega \\
&= |\mathbf{I}_d - 2t\mathbf{P}\sigma^2\mathbf{I}_d|^{-\frac{1}{2}} \\
&= \left(\prod_{i=1}^{\text{rang}(\mathbf{P})}(1 - 2\sigma^2 t\lambda_i)\right)^{-\frac{1}{2}}
\end{aligned}$$

Da der yderligere gælder $M_{X^T\mathbf{P}X}(t) = (1 - 2\sigma^2 t)^{-\frac{r}{2}}$, når $X^T\mathbf{P}X \sim \sigma^2\chi_r^2$, får man

$$(1 - 2\sigma^2 t)^{-\frac{r}{2}} = \left(\prod_{i=1}^{\text{rang}(\mathbf{P})}(1 - 2\sigma^2 t\lambda_i)\right)^{-\frac{1}{2}}$$

Dermed ses, at $\lambda_i = 1$, når $i = 1, \cdots, \text{rang}(\mathbf{P})$, hvor $\text{rang}(\mathbf{P}) = r$. Derfor er $\mathbf{P}$ en symmetrisk $d \times d$ matrix med r egenværdier med værdien er 1 og $d - r$ egenværdier med værdien 0. Heraf følger fra sætning A.0.3, at $\mathbf{P}$ er idempotent med rang r.

Hvis det omvendt antages, at $\mathbf{P}^2 = \mathbf{P}$ og $\text{rang}(\mathbf{P}) = r$, findes der en ortogonal matrix $\mathbf{T}$ sådan, at

$$\mathbf{T}^T\mathbf{P}\mathbf{T} = \Lambda = \begin{pmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{pmatrix}$$

Da det er givet, at $X \sim N_d(0, \sigma^2\mathbf{I}_d)$, får man

$$\begin{aligned}
\begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X \sim N_r &\left(0, \quad \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}\sigma^2\mathbf{I}_d\mathbf{P}\mathbf{T} \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix}\right) \\
= N_r &\left(0, \quad \sigma^2 \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}\mathbf{I}_d\mathbf{P}\mathbf{T} \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix}\right) \\
= N_r &\left(0, \quad \sigma^2 \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}^2\mathbf{T} \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix}\right) \\
= N_r &\left(0, \quad \sigma^2 \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}\mathbf{T} \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix}\right) \\
= N_r &\left(0, \quad \sigma^2 \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \Lambda \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix}\right) \\
= N_r &\left(0, \sigma^2\mathbf{I}_r\right).
\end{aligned}$$

Der gælder altså at $\left(\begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X\right)^T \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X \sim \sigma^2\chi^2_r$, hvliket kan omskrives til

$$\begin{aligned}
\left(\begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X\right)^T \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X &= X^T\mathbf{P}\mathbf{T} \begin{pmatrix} \mathbf{I}_r \\ 0 \end{pmatrix} \begin{pmatrix} \mathbf{I}_r & 0 \end{pmatrix} \mathbf{T}^T\mathbf{P}X \\
&= X^T\mathbf{P}\mathbf{T}\Lambda\mathbf{T}^T\mathbf{P}X \\
&= X^T\mathbf{P}\mathbf{P}\mathbf{P}X \\
&= X^T\mathbf{P}X,
\end{aligned}$$

som viser at $X^T\mathbf{P}X \sim \sigma^2\chi^2_r$. ■

Sætning A.0.6. *Lad* $X \sim N_d(0, \sigma^2\boldsymbol{I}_d)$ *og* $Q_i = \dfrac{X^T\boldsymbol{P}_iX}{\sigma^2}$ *være fordelt som* $\chi^2_{r_i}$*, hvor* $i = 1, 2$*. Da er* Q_1 *og* Q_2 *uafhængige, hvis og kun hvis* $\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{O}$*.*

Bevis
antag at Q_1 og Q_2 er uafhængige, da gælder, fordi summen af uafhængige chi-i-anden fordelte variable er chi-i-anden fordelte, at $\dfrac{X^T(\mathbf{P}_1 + \mathbf{P}_2)X}{\sigma^2}$ er fordelt som $\chi^2_{r_1+r_2}$. Dermed har man af sætning A.0.5, at $\mathbf{P}_1, \mathbf{P}_2$ og $\mathbf{P}_1 + \mathbf{P}_2$ er idempotente, sådan at

$$\begin{aligned}
(\mathbf{P}_1 + \mathbf{P}_2)^2 &= \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_2\mathbf{P}_1 \\
&= \mathbf{P}_1 + \mathbf{P}_2
\end{aligned}$$

hvilket viser at $\mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_2\mathbf{P}_1 = \mathbf{O}$. Ved at højre- og venstre multiplicere denne ligning med $\mathbf{P}_1$ får man

$$\begin{aligned}
&\mathbf{P}_1\mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_1\mathbf{P}_2\mathbf{P}_1 = \mathbf{O}, \\
&\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1 + \mathbf{P}_2\mathbf{P}_1\mathbf{P}_1 = \mathbf{O}.
\end{aligned}$$

Det følger, da både $\mathbf{P}_1$ og $\mathbf{P}_2$ er idempotente, at

$$\begin{aligned}\mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_1\mathbf{P}_2\mathbf{P}_1 &= \mathbf{P}_1\mathbf{P}_2\mathbf{P}_1 + \mathbf{P}_2\mathbf{P}_1 \\ \mathbf{P}_1\mathbf{P}_2 &= \mathbf{P}_2\mathbf{P}_1 \\ \mathbf{P}_1\mathbf{P}_2 - \mathbf{P}_2\mathbf{P}_1 &= \mathbf{O}\end{aligned}$$

Da både $\mathbf{P}_1\mathbf{P}_2 - \mathbf{P}_2\mathbf{P}_1 = \mathbf{O}$ og $\mathbf{P}_1\mathbf{P}_2 + \mathbf{P}_2\mathbf{P}_1 = \mathbf{O}$ gælder, at $\mathbf{P}_1\mathbf{P}_2 = \mathbf{O}$
Antages omvendt, at $\mathbf{P}_1\mathbf{P}_2 = \mathbf{O}$, så gælder

$$\begin{aligned}Cov[\mathbf{P}_1X, \mathbf{P}_2X] &= \mathbf{P}_1 Cov[X,X]\mathbf{P}_2^T \\ &= \mathbf{P}_1 Var[X]\mathbf{P}_2^T \\ &= \sigma^2\mathbf{P}_1\mathbf{P}_2^T \\ &= \sigma^2\mathbf{P}_1\mathbf{P}_2 \\ &= \mathbf{O}.\end{aligned}$$

Hermed følger af sætning 4.2.4, at $\mathbf{P}_1X$ og $\mathbf{P}_2X$ er uafhængige. Dermed er også Q_i, $i = 1, 2$, uafhængige, da

$$||\mathbf{P}_iX||^2 = (\mathbf{P}_iX)^T(\mathbf{P}_iX) = X^T\mathbf{P}_i^2X = X^T\mathbf{P}_iX = \sigma^2 Q_i.$$

■

Sætning A.0.7.

$$\begin{vmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{vmatrix} = \begin{cases} |\boldsymbol{D}||\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}|, & \textit{hvis } \boldsymbol{D} \textit{ er regulær,} \\ |\boldsymbol{A}||\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}|, & \textit{hvis } \boldsymbol{A} \textit{ er regulær.} \end{cases}$$

Bevis

Antag, at $\mathbf{D}$ er regulær, da gælder

$$\begin{vmatrix} \mathbf{I}_r & \mathbf{O} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_s \end{vmatrix} = |\mathbf{I}_r||\mathbf{I}_s| = 1.$$

Heraf følger

$$\begin{aligned}\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} &= \left|\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\begin{pmatrix} \mathbf{I}_r & \mathbf{O} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_s \end{pmatrix}\right| \\ &= \begin{vmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{B} \\ \mathbf{C} - \mathbf{D}\mathbf{D}^{-1}\mathbf{C} & \mathbf{D} \end{vmatrix} \\ &= \begin{vmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} \end{vmatrix} \\ &= |\mathbf{D}||\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}|.\end{aligned}$$

Antag nu istedet, at $\mathbf{A}$ er regulær, da gælder

$$\begin{vmatrix} \mathbf{I}_r & -\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{O} & \mathbf{I}_s \end{vmatrix} = |\mathbf{I}_r||\mathbf{I}_s| = 1.$$

Heraf følger

$$\begin{aligned}\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} &= \left|\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & -\mathbf{A}^{-1}\mathbf{B}\\ \mathbf{O} & \mathbf{I}_s\end{pmatrix}\right| \\ &= \begin{vmatrix}\mathbf{A} & -\mathbf{A}\mathbf{A}^{-1}\mathbf{B}+\mathbf{B}\\ \mathbf{C} & -\mathbf{C}\mathbf{A}^{-1}\mathbf{B}+\mathbf{D}\end{vmatrix} \\ &= \begin{vmatrix}\mathbf{A} & \mathbf{O}\\ \mathbf{C} & \mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\end{vmatrix} \\ &= |\mathbf{A}||\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|.\end{aligned}$$

■

Sætning A.0.8. *Hvis $\boldsymbol{A} > \boldsymbol{O}, \boldsymbol{B} \geq \boldsymbol{O}$ samt $c > 0$ og $d > 0$ gælder $c\boldsymbol{A} + d\boldsymbol{B} > \boldsymbol{O}$.*

Bevis

Betragt udtrykket $x^T(c\mathbf{A} + d\mathbf{B})x$. Da der gælder at $x^T\mathbf{A}x > 0$ og $x^T\mathbf{B}x \geq 0$ følger

$$x^T(c\mathbf{A} + d\mathbf{B})x = cx^T\mathbf{A}x + dx^T\mathbf{B}x > 0$$

Hvilket viser at matricen $c\mathbf{A} + d\mathbf{B}$ er positiv definit. ■

Sætning A.0.9. *Lad $\boldsymbol{E}$ og $\boldsymbol{H}$ være $d \times d$ matricer, som er positiv definite, og lad $\boldsymbol{Z} = \boldsymbol{E} + \boldsymbol{H}$ og $\boldsymbol{V} = (\boldsymbol{E}+\boldsymbol{H})^{-\frac{1}{2}}\boldsymbol{H}(\boldsymbol{E}+\boldsymbol{H})^{-\frac{1}{2}}$, da gælder*

$$\frac{\partial(\boldsymbol{H},\boldsymbol{E})}{\partial(\boldsymbol{V},\boldsymbol{Z})} = |\boldsymbol{Z}|^{\frac{d+1}{2}}.$$

Bevis

Da Jacobimatricer er er multiplikative, kan følgende opstilles:

$$\frac{\partial(\mathbf{V},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{E})} = \frac{\partial(\mathbf{V},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{Z})}\frac{\partial(\mathbf{H},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{E})}$$

Ideen er nu at finde de to Jacobideterminanter på højre side af lighedstegnet. Sættes $\mathbf{W} = \mathbf{Z}$ fås, ved differentiation med hensyn til $\mathbf{H}$, da $\mathbf{W} = \mathbf{Z}$ her betragtes som en konstant, at

$$\frac{\partial w_i}{\partial h_j} = 0 \ \text{ og } \ \frac{\partial w_i}{\partial z_j} = \delta_{ij}, \ \text{ hvor } \delta_{ij} \text{ er Kroneckers delta.}$$

Det følger nu fra definitionen af Jacobimatricen, at

$$\frac{\partial(\mathbf{V},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{Z})} = \frac{\partial(\mathbf{V},\mathbf{W})}{\partial(\mathbf{H},\mathbf{Z})} = \begin{vmatrix}\frac{\partial\mathbf{V}}{\partial\mathbf{H}} & \frac{\partial\mathbf{W}}{\partial\mathbf{H}}\\ \frac{\partial\mathbf{V}}{\partial\mathbf{Z}} & \frac{\partial\mathbf{W}}{\partial\mathbf{Z}}\end{vmatrix} = \begin{vmatrix}\frac{\partial\mathbf{V}}{\partial\mathbf{H}} & \mathbf{O}\\ \frac{\partial\mathbf{V}}{\partial\mathbf{Z}} & \mathbf{I}\end{vmatrix} = \left|\frac{\partial\mathbf{V}}{\partial\mathbf{H}}\right|$$

hvilket er Jacobideterminanten til $\dfrac{\partial\mathbf{V}}{\partial\mathbf{H}}$.
Da $\mathbf{E}$ og $\mathbf{H}$ er positiv definite, gælder ifølge sætning A.0.8, at $\mathbf{Z}$ er positiv definit. Yderligere, da $\mathbf{Z}$ er symmetrisk, gælder, at $\mathbf{V} = \mathbf{Z}^{-\frac{1}{2}}\mathbf{H}(\mathbf{Z}^{-\frac{1}{2}})^T$.
Det udnyttes nu, at det kan vises, at hvis $\mathbf{X}$ og $\mathbf{Y}$ er $d \times d$ matricer, og $\mathbf{A}$ er regulær, samt at $\mathbf{X} = \mathbf{A}\mathbf{Y}\mathbf{A}^T$, gælder $\dfrac{\partial\mathbf{X}}{\partial\mathbf{Y}} = |\mathbf{A}|^{d+1}$.

Specielt gælder, hvis $\mathbf{X} = \mathbf{A}^{\frac{1}{2}}\mathbf{Y}(\mathbf{A}^{\frac{1}{2}})^T$ at $\dfrac{\partial \mathbf{X}}{\partial \mathbf{Y}} = |\mathbf{A}|^{\frac{d+1}{2}}$.

Her får vi, at $\dfrac{\partial \mathbf{V}}{\partial \mathbf{H}} = |\mathbf{Z}|^{\frac{-(d+1)}{2}}$. Der differentieres med hensyn til $\mathbf{E}$, hvor $\mathbf{H}$ betragtes som en konstant og

$$\frac{\partial w_i}{\partial e_j} = \frac{\partial z_i}{\partial e_j} = \frac{\partial e_i}{\partial e_j} + \frac{\partial h_i}{\partial e_j} = \delta_{ij} + 0.$$

Man får da

$$\frac{\partial(\mathbf{H},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{E})} = \frac{\partial(\mathbf{H},\mathbf{W})}{\partial(\mathbf{H},\mathbf{E})} = \begin{vmatrix} \dfrac{\partial \mathbf{H}}{\partial \mathbf{H}} & \dfrac{\partial \mathbf{W}}{\partial \mathbf{H}} \\ \dfrac{\partial \mathbf{H}}{\partial \mathbf{E}} & \dfrac{\partial \mathbf{W}}{\partial \mathbf{E}} \end{vmatrix} = \begin{vmatrix} \mathbf{I} & \mathbf{O} \\ \dfrac{\partial \mathbf{H}}{\partial \mathbf{E}} & \mathbf{I} \end{vmatrix} = 1.$$

Det følger at

$$\frac{\partial(\mathbf{H},\mathbf{E})}{\partial(\mathbf{V},\mathbf{Z})} = \left(\frac{\partial(\mathbf{V},\mathbf{Z})}{\partial(\mathbf{H},\mathbf{E})}\right)^{-1} = \left(|\mathbf{Z}|^{\frac{-(d+1)}{2}} \cdot 1\right)^{-1} = |\mathbf{Z}|^{\frac{d+1}{2}}.$$

■