Kapitel 6

FORDELINGER PÅ REELLE TALRUM

6.1. Tætheder og fordelinger.

Flerdimensionale kontinuerte fordelinger er ikke sværere at forstå end de éndimensionale. Men rent teknisk dukker der nye aspekter op, bl.a. fordi der er flere muligheder for transformation af sådanne fordelinger.

EKSEMPEL 6.1.1 (ligefordelingen på enhedskvadratet). Betragt en todimensional stokastisk variabel

$$X = (X_1, X_2) \in \mathbf{R}^2$$

med følgende opførsel. Der gælder $P(0 \leq X_1 \leq 1, 0 \leq X_2 \leq 1) = 1$, dvs. hele sandsynlighedsmassen er placeret i enhedskvadratet $[0,1] \times [0,1]$. For enhver pæn delmængde A af enhedskvadratet gælder

$$P(X \in A) = \text{arealet af } A.$$

Ved en pæn mængde forstår vi her det der i matematisk analyse ofte kaldes et "område", altså for eksempel et rektangel, en foreningsmængde af endeligt mange rektangler, en polygon, en cirkelskive; kort sagt, enhver mængde der med rimelighed kan tænkes at udspringe af en praktisk problemstilling, og ikke netop er konstrueret af skrupelløse matematikere med det formål at fungere som modeksempel på et eller andet.

For $0 \leq a_1 \leq b_1 \leq 1$ og $0 \leq a_2 \leq b_2 \leq 1$ kan vi specielt betragte mængden $A = [a_1, b_1] \times [a_2, b_2]$, som er et rektangel med sider der er parallelle med akserne. Da arealet af dette er $(b_1 - a_1)(b_2 - a_2)$ får vi

$$P(X_1 \in [a_1, b_1], X_2 \in [a_2, b_2]) = (b_1 - a_1)(b_2 - a_2).$$

Specielt fås for $a_1 = 0$, $b_1 = 1$,

$$P(X_2 \in [a_2, b_2]) = b_2 - a_2,$$

dvs. den marginale fordeling af X_2 er ligefordelingen på enhedsintervallet. Tilsvarende ses at fordelingen af X_1 er ligefordelingen, og da der således gælder

$$P(X_1 \in [a_1, b_1], X_2 \in [a_2, b_2]) = P(X_1 \in [a_1, b_1])P(X_2 \in [a_2, b_2])$$

er det naturligt at opfatte X_1 og X_2 som stokastisk uafhængige.

Tilsvarende kan man i højere dimension definere ligefordelingen på "enhedsterningen" $[0,1]^n$ som den fordeling, der til en pæn mængde tilordner dens rumfang ($n=3$) eller n-dimensionale volumen ($n \geq 4$). En stokastisk variabel med denne fordeling kan også fortolkes som et sæt bestående af n stokastisk uafhængige, ligefordelte variable på enhedsintervallet. Disse fordelinger siges at have tæthed 1 på enhedskvadratet (–terningen), da forholdet mellem sandsynlighed og areal (rumfang, n–dimensionalt volumen) for enhver delmængde netop er 1.

Integration på $\mathbf{R}^n$. For at komme videre er vi nødt til at beskæftige os lidt med integration af funktioner i flere variable. For at lette notationen ser vi kun på tilfældet $n=2$, og nøjes med at betragte en kontinuert funktion $f\colon [0,1]^2 \to \mathbf{R}$. *Integralet* (eller *planintegralet*) af en sådan funktion kan defineres som grænseværdi for middelsummer, f.eks. ved

$$\int_0^1 \int_0^1 f(x_1,x_2)\,dx_1\,dx_2 = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} f\left(\frac{i-0.5}{n}, \frac{j-0.5}{n}\right).$$

Her kan det enkelte led $\frac{1}{n^2} f\left(\frac{i-0.5}{n}, \frac{j-0.5}{n}\right)$ fortolkes som værdien af f i et punkt (i dette tilfælde midtpunktet) i kvadratet $[\frac{i-1}{n}, \frac{i}{n}] \times [\frac{j-1}{n}, \frac{j}{n}]$ med sidelængde $\frac{1}{n}$, ganget med arealet $\frac{1}{n^2}$ af dette kvadrat. Middelsummen fremkommer altså ved summation af disse bidrag over alle kvadrater i en (finere og finere) inddeling af integrationsområdet i småkvadrater. Beviset for, at konvergensen finder sted, er ikke meget forskelligt fra beviset for det tilsvarende resultat i dimension 1.

For fastholdt i kan summen

$$\frac{1}{n} \sum_{j=1}^{n} f\left(\frac{i-0.5}{n}, \frac{j-0.5}{n}\right)$$

opfattes som en middelsum, der approksimerer det sædvanlige éndimensionale integral

$$\int_0^1 f\left(\frac{i-0.5}{n}, x_2\right) dx_2.$$

Lad nu $f_1\colon [0,1] \to \mathbf{R}$ betegne den funktion der fremkommer af f når x_2 "integreres ud", altså

$$f_1(x_1) = \int_0^1 f(x_1,x_2)\,dx_2.$$

Så er, ifølge ovenstående,

$$\begin{aligned} &\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} f\left(\frac{i-0.5}{n}, \frac{j-0.5}{n}\right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} f\left(\frac{i-0.5}{n}, \frac{j-0.5}{n}\right)\right) \approx \frac{1}{n} \sum_{i=1}^{n} f_1\left(\frac{i-0.5}{n}\right). \end{aligned}$$

Men det sidste udtryk er igen en middelsum, der approksimerer integralet af den kontinuerte funktion f_1. I grænsen fås derfor identiteten

$$\begin{aligned}&\int_0^1 \int_0^1 f(x_1, x_2)\, dx_1\, dx_2 \\ &= \int_0^1 f_1(x_1)\, dx_1 = \int_0^1 \left(\int_0^1 f(x_1, x_2)\, dx_2 \right) dx_1\end{aligned}$$

som udtrykker, at et planintegral kan udregnes ved en dobbelt integration, hvor først den ene og så den anden af de to variable integreres ud (og rækkefølgen er i øvrigt ligegyldig).

På tilsvarende måde defineres integraler på $\mathbf{R}^n$, og der gælder den tilsvarende regel om udregning ved "n–dobbelt" integration m.h.t. de enkelte variable.

Mere vil vi ikke sige om integration af funktioner i flere variable. Der tilbagestår adskillige udvidelser, f.eks. til stykkevis kontinuerte funktioner, funktioner som er udefinerede i visse punkter eller på visse flader af lavere dimension, funktioner som er definerede på et ubegrænset område etc. etc. Men disse udvidelser er analoge til dem vi kender fra dimension 1.

Eksempel 6.1.2. Lad A være et pænt, begrænset område i planen. Da er

$$\int\int 1_A(x_1, x_2)\, dx_1\, dx_2$$

veldefineret som grænseværdi for middelsummer svarende til finere og finere kvadreringer af planen, og grænseværdien bliver netop arealet af området A. Dette følger af, at en middelsum vil være sum af arealerne af nogle af de små kvadrater. Summen vil omfatte alle småkvadrater som er helt indeholdt i A og ingen af dem der ligger helt udenfor A. Småkvadrater som skæres af A's rand kan enten være med eller ikke, det er ligegyldigt, fordi disses arealer repræsenterer et forsvindende bidrag i grænsen. Vi får altså netop det samlede areal af A som grænseværdi, og hermed har vi i øvrigt blot gentaget en intuitivt let forståelig og hyppigt benyttet definition af et områdes areal.

Tilsvarende vil

$$\int\int\int 1_A(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3$$

for en pæn begrænset delmængde af $\mathbf{R}^3$ (f.eks. en kugle) kunne fortolkes som A's rumfang, og i højere dimensioner end 3 kan vi opfatte det tilsvarende integral som en definition af begrebet n–dimensionalt volumen.

Definition. Ved en *sandsynlighedstæthed* på $\mathbf{R}^n$ forstår vi en ikke–negativ funktion $p\colon \mathbf{R}^n \to \mathbf{R}$ med

$$\int \cdots \int p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 1.$$

Den til tætheden p hørende *sandsynlighedsfordeling* er afbildningen, som til en pæn delmængde A af $\mathbf{R}^n$ tilordner tallet

$$P(A) = \int \cdots \int 1_A(x_1, \ldots, x_n) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

Hermed har vi indført de kontinuerte fordelinger i flere dimensioner. Notationen i forbindelse med stokastiske variable kan overføres uændret fra det diskrete og det éndimensionalt kontinuerte tilfælde, så den vil vi ikke sige noget om. Der mangler naturligvis en afklaring af, i hvilket omfang og for hvilke transformationer t en afledt variabel $t(X)$ kan tillægges en fordeling af én af de typer vi har indført, men det vender vi tilbage til i næste paragraf.

Den intuitive fortolkning af tætheden p er analog til den vi kender fra én dimension. Blot skal "et lille interval omkring eller nær ved x" erstattes med "en lille mængde A omkring eller nær ved $x = (x_1, \ldots, x_n)$". Denne mængde kan for eksempel være en lille cirkelskive (kugle) med centrum i x eller et kvadrat (en terning) af formen

$$A = [x_1, x_1 + h] \times \ldots \times [x_n, x_n + h]$$

for et lille tal h. Hvis x er et kontinuitetspunkt for tætheden p kan vi for et givet $\epsilon > 0$, ved at vælge A tilstrækkeligt lille, opnå at der for alle $x' \in A$ gælder

$$p(x) - \epsilon \leq p(x') \leq p(x) + \epsilon.$$

Så fås, ved integration af denne ulighed over mængden A, idet konstanterne $p(x) \pm \epsilon$ sættes udenfor integraltegnet og A's areal (rumfang, n-dimensionale volumen) betegnes $|A|$,

$$(p(x) - \epsilon)|A| \leq \int \cdots \int 1_A(x'_1, \ldots, x'_n) p(x'_1, \ldots, x'_n)\, dx'_1, \ldots,\, dx'_n \leq (p(x) + \epsilon)|A|.$$

Men det betyder netop, at der for en stokastisk variabel X med denne tæthed vil gælde

$$P(X \in A) \approx p(x)|A|,$$

dvs. tætheden er det lokale forhold mellem sandsynlighed og areal (rumfang, n-dimensionalt volumen).

Eksempel 6.1.3. Lad p_1 og p_2 være sandsynlighedstætheder på $\mathbf{R}$, og betragt funktionen $p\colon \mathbf{R}^2 \to \mathbf{R}$ defineret ved

$$p(x_1, x_2) = p_1(x_1) p_2(x_2).$$

Denne funktion er en sandsynlighedstæthed, idet

$$\begin{aligned}\int\int p(x_1,x_2)\,dx_1\,dx_2 &= \int\left(\int p_1(x_1)p_2(x_2)\,dx_1\right)dx_2\\ = \int p_2(x_2)\left(\int p_1(x_1)\,dx_1\right)dx_2 &= \int p_2(x_2)\,dx_2 = 1.\end{aligned}$$

Ved analogi med det diskrete tilfælde (se f.eks. sætning 2.4.1) er det rimeligt at opfatte denne fordeling som fordelingen af (X_1, X_2), hvor X_1 og X_2 er stokastiske uafhængige reelle variable med tætheder p_1 og p_2. Dette understøttes af, at vi for delmængder A_1 og A_2 af $\mathbf{R}$ (f.eks. intervaller) har

$$\begin{aligned}P(X_1\in A_1) &= \int\int 1_{A_1\times\mathbf{R}}(x_1,x_2)p_1(x_1)p_2(x_2)\,dx_1\,dx_2\\ &= \int p_2(x_2)\left(\int 1_{A_1}(x_1)p_1(x_1)\,dx_1\right)dx_2 = \int 1_{A_1}(x_1)p_1(x_1)\,dx_1\end{aligned}$$

og tilsvarende

$$P(X_2\in A_2) = \int 1_{A_2}(x_2)p_2(x_2)\,dx_2,$$

hvilket betyder at X_1 og X_2 kan opfattes som kontinuert fordelte éndimensionale variable med de givne tætheder p_1 og p_2. Videre er

$$\begin{aligned}&P(X_1\in A_1 \text{ og } X_2\in A_2)\\ &= \int\int 1_{A_1\times A_2}(x_1,x_2)p_1(x_1)p_2(x_2)\,dx_1\,dx_2\\ &= \int\left(\int 1_{A_1}(x_1)1_{A_2}(x_2)p_1(x_1)p_2(x_2)\,dx_1\right)dx_2\\ &= \int 1_{A_2}(x_2)p_2(x_2)\left(\int 1_{A_1}(x_1)p_1(x_1)\,dx_1\right)dx_2\\ &= P(X_1\in A_1)\int 1_{A_2}(x_2)p_2(x_2)\,dx_2\\ &= P(X_1\in A_1)P(X_2\in A_2),\end{aligned}$$

i overensstemmelse med definitionen af uafhængighed i det diskrete tilfælde, jvf. §2.4.

Eksemplet inspirerer således til følgende

DEFINITION. Lad $(X_1,\dots,X_n)$ følge fordelingen med tæthed p på $\mathbf{R}^n$. $X_1,\dots,X_n$ siges at være *stokastisk uafhængige* hvis tætheden kan skrives på formen

$$p(x_1,\dots,x_n) = p_1(x_1)\dots p_n(x_n)$$

hvor $p_1, \ldots, p_n$ er sandsynlighedstætheder på $\mathbf{R}$.

Fortolkningen af $p_1, \ldots, p_n$ er naturligvis, at disse funktioner er tæthederne for de marginale fordelinger af de n variable $X_1, \ldots, X_n$. Dette vil blive formuleret lidt mere præcist i næste paragraf. Den regneregel, vi i eksempel 6.1.3 udledte for $n = 2$, generaliserer umiddelbart til

$$P\left((X_1, \ldots, X_n) \in A_1 \times \ldots \times A_n\right) = P\left(X_1 \in A_1\right) \ldots P\left(X_n \in A_n\right).$$

Middelværdi, varians og kovarians. I næste paragraf skal vi i detaljer studere det problem, der består i for en given transformation $t\colon \mathbf{R}^n \to E$ (hvor E specielt kan være den reelle akse) og en given tæthed p på $\mathbf{R}^n$ at bestemme den fordeling på E der med rette kan opfattes som fordelingen af den afledte variable $Y = t(X)$, når X's fordeling er givet ved tætheden p. Det krav, vi vil stille til fordelingen af Y, er naturligvis at der skal gælde

$$\begin{aligned} P(Y \in B) &= P\left(X \in t^{-1}(B)\right) \\ &= \int \cdots \int 1_{t^{-1}(B)}(x_1, \ldots, x_n) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n. \end{aligned}$$

Denne betingelse kan også skrives

$$\mathrm{E}1_B(Y) = \int \cdots \int 1_B(t(x_1, \ldots, x_n)) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

hvor venstre side defineres ved summation eller integration over E, afhængigt af om fordelingen af Y er diskret eller kontinuert. Ved at danne linearkombinationer af sådanne relationer får man umiddelbart, at identiteten

$$\mathrm{E}f(Y) = \int \cdots \int f(t(x_1, \ldots, x_n)) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

vil være gyldig for enhver funktion f på E som er stykkevis konstant. Herfra er der kun et lille skridt til at indse, at identiteten er gyldig for alle funktioner f, i den forstand at summen/integralet på venstre side er defineret hvis og kun hvis integralet på højre side er defineret, og i så fald er de ens. Beviset for dette – som strengt taget burde gennemføres for hver enkelt af de typer af transformationsproblemer, vi studerer i næste paragraf – er i alle tilfælde helt elementært. Vi vil ikke forplumre begreberne ved ustandselig at trække problemer af denne art frem. Derfor vil vi nøjes med her at konstatere, at når vi definerer fordelinger af afledte variable på en sådan måde, at sandsynligheder opfører sig konsistent (som beskrevet ovenfor), så vil middelværdier åbenbart også gøre det. Specielt vil det gælde at middelværdien af en afledt reel variabel $f(X_1, \ldots, X_n)$

frit kan beregnes enten i sin egen (éndimensionale) fordeling på sædvanlig måde, eller ved integration i fordelingen af $(X_1, \ldots, X_n)$ efter formlen

$$\mathrm{E}f(X_1, \ldots, X_n) = \int \cdots \int f(x_1, \ldots, x_n) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

Alle definitioner og resultater fra kapitel 4 kan overføres stort set uændrede, med den gennemgående forskel at summer skal erstattes med integraler. Som eksempel på dette kan vi se på beviset for (c) i sætning 4.2.2. Påstanden dér er, at hvis $Y_1, \ldots, Y_n$ er stokastisk uafhængige vil der gælde $\mathrm{E}(Y_1 \ldots Y_n) = (\mathrm{E}Y_1) \ldots (\mathrm{E}Y_n)$. Hvis man i beviset, som kun er gennemført for $n = 2$, erstatter alle summer med tilsvarende integraler (idet den første dobbeltsum skal erstattes med et planintegral) får man umiddelbart et bevis for samme påstands gyldighed i det kontinuerte tilfælde. Også definitionerne af kovarians og korrelation kan overføres uden videre, og der gælder de samme regneregler som i det diskrete tilfælde. Chebychev's ulighed gælder uforandret for uafhængige variable med kontinuerte fordelinger, de store tals lov ligeså.

6.2. Transformation.

I slutningen af §5.3 (i eksempel 5.3.4) konstaterede vi, at det ikke altid er muligt, inden for de rammer vi har afstukket, at tillægge en transformeret variabel $Y = t(X)$ en fordeling. Dér drejede det sig om en situation, man ville kunne løse ved at udvide fordelingesbegrebet en smule, så det kom til at omfatte fordelinger af blandet diskret–kontinuert type. I flere dimensioner bliver det en tand værre, som følgende eksempel illustrerer:

Eksempel 6.2.1. Lad X være ligefordelt på enhedsintervallet, og betragt den todimensionale afledte variable (X, X^2). Fordelingen af denne variabel er åbenbart koncentreret på et stykke af en parabel. Det er ikke nogen diskret fordeling, og den kan heller ikke beskrives ved en tæthed i planen.

Eksemplet viser, at man let kan komme til at opskrive variable, der ikke kan tillægges en fordeling af en af de typer vi har indført. Samtidig skulle det gerne fremgå, at dette – i hvert fald i nogle tilfælde – er lidt af et pseudoproblem. Hvis man virkelig fik brug for at udregne sandsynligheden for, at (X, X^2) havnede i en given mængde, ville det jo være let at føre problemet tilbage til en hændelse der vedrørte X. I praksis kan næsten alle transformationsproblemer løses ved kombination af de tre hovedtilfælde, vi nu vil gennemgå.

Tilfælde 1: Diskret afledt variabel. Lad $t\colon \mathbf{R}^n \to E$ være stykkevis konstant, E en vilkårlig mængde (typisk $\mathbf{N}$ eller $\mathbf{Z}$). Her underforstås at "stykkerne" $t^{-1}(y), y \in E$, er pæne mængder. For en stokastisk variabel X med tæthed p på $\mathbf{R}^n$ ligger det lige for at definere fordelingen af

$Y = t(X)$ som den diskrete fordeling med punktsandsynligheder

$$\begin{aligned} P(Y = y) &= P\left(X \in t^{-1}(y)\right) \\ &= \int \cdots \int 1_{t^{-1}(y)}(x_1, \ldots, x_n) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n. \end{aligned}$$

Situationen adskiller sig ikke væsentligt fra den tilsvarende i dimension 1, så vi vil ikke sige mere om den.

Tilfælde 2: Marginal fordeling. Lad, for $k < n$, $t\colon \mathbf{R}^n \to \mathbf{R}^k$ være projektionen ned på k af de n koordinatakser. For nemheds skyld kan vi antage at det drejer sig om de k første, således at t simpelthen er den afbildning der "glemmer" de sidste $n - k$ koordinater, altså

$$t(x_1, \ldots, x_n) = (x_1, \ldots, x_k).$$

Med andre ord, for en n-dimensional variabel $X = (X_1, \ldots, X_n)$ med tæthed p ønsker vi at give en fornuftig definition af fordelingen af sættet $Y = (X_1, \ldots, X_k)$, bestående af de k første koordinater.

SÆTNING 6.2.1. *Definer $q\colon \mathbf{R}^k \to \mathbf{R}$ ved*

$$q(x_1, \ldots, x_k) = \int \cdots \int p(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n.$$

For enhver pæn mængde $B \subseteq \mathbf{R}^k$ gælder da

$$\begin{aligned} &\int \cdots \int 1_{t^{-1}(B)}(x_1, \ldots, x_n) p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\ &= \int \cdots \int 1_B(x_1, \ldots, x_k) q(x_1, \ldots, x_k)\, dx_1 \ldots dx_k. \end{aligned}$$

BEVIS. Vi nøjes med at opskrive beviset i tilfældet $n = 2$, $k = 1$. Beviset er i princippet ikke sværere i det generelle tilfælde, men med den notation vi benytter her er det så ufatteligt besværligt at opskrive, at den dybere mening med garanti ville gå tabt hvis vi forsøgte.

Vi ser altså på en tæthed p i planen, og $t\colon \mathbf{R}^2 \to \mathbf{R}$ er projektionen ned på første akse, givet ved $t(x_1, x_2) = x_1$. Mængden $B \subseteq \mathbf{R}$ kan vi for nemheds skyld tænke på som et interval. Så er $t^{-1}(B) = B \times \mathbf{R}$ en lodret strimmel, og funktionen $q\colon \mathbf{R} \to \mathbf{R}$ er givet ved $q(x_1) = \int p(x_1, x_2)\, dx_2$. Vi får

$$\begin{aligned} &\int \int 1_{t^{-1}(B)}(x_1, x_2) p(x_1, x_2)\, dx_1\, dx_2 \\ &= \int \left(\int 1_B(x_1) 1_{\mathbf{R}}(x_2) p(x_1, x_2)\, dx_2 \right) dx_1 \\ &= \int 1_B(x_1) \left(\int p(x_1, x_2)\, dx_2 \right) dx_1 \\ &= \int 1_B(x_1) q(x_1)\, dx_1, \end{aligned}$$

hvilket skulle vises.

Herefter er det klart, at vi må definere fordelingen af $Y = (X_1, \ldots, X_k)$ som den kontinuerte fordeling på $\mathbf{R}^k$ med tæthed q; at q er en sandsynlighedstæthed følger jo umiddelbart af sætningen for $B = \mathbf{R}^k$. Indholdet af sætningen og denne definition kan i alt væsentligt sammenfattes således:

I en flerdimensional fordeling med tæthed fremkommer tætheden for den marginale fordeling af nogle af variablene ved, at man integrerer de øvrige variable ud.

Eksempel 6.2.2. Lad (X_1, X_2) være ligefordelt på området

$$K = \{(x_1, x_2) \mid |x_1| + |x_2| \leq 1\},$$

som ses at være et kvadrat med sidelængde $\sqrt{2}$ og centrum $(0, 0)$, roteret 45^o i forhold til koordinatakserne. Ved ligefordelingen på denne mængde forstår vi naturligvis fordelingen med tæthed $p(x_1, x_2) = \frac{1}{2} 1_K(x_1, x_2)$, hvor normeringsfaktoren $\frac{1}{2}$ skyldes at K har areal 2. Fordelingen af X_1 har da tætheden

$$q(x_1) = \frac{1}{2} \int 1_K(x_1, x_2)\, dx_2 = \frac{1}{2} \int_{-(1-|x_1|)}^{+(1-|x_1|)} 1\, dx_2 = 1 - |x_1|$$

på intervallet $-1 \leq x_1 \leq 1$.

Eksempel 6.2.3. Lad $(X_1, \ldots, X_n)$ have tæthed af formen

$$p(x_1, \ldots, x_n) = p_1(x_1) \ldots p_n(x_n),$$

hvor $p_1, \ldots, p_n$ er givne sandsynlighedstætheder på den reelle akse. Det følger da let af sætning 6.2.1 og den efterfølgende definition, at fordelingen af den k–dimensionale variable $(X_1, \ldots, X_k)$ har tætheden

$$q(x_1, \ldots, x_k) = p_1(x_1) \ldots p_k(x_k).$$

Specielt ses at den marginale fordeling af X_1 (eller X_i) får tætheden p_1 (eller p_i). Efter definitionen af uafhængighed sidst i §6.1 kan vi altså sammenfatte denne situation ved at sige, at $X_1, \ldots, X_n$ er uafhængige, reelle variable med tætheder $p_1, \ldots, p_n$.

Tilfælde 3: Dimensionsbevarende transformation. Betragt en n-dimensional stokastisk variabel $X = (X_1, \ldots, X_n)$ med tæthed p, og lad $t\colon \mathbf{R}^n \to \mathbf{R}^n$ være en given transformation. Under passende regularitetsbetingelser på t er det rimeligt at gætte på, at $Y = t(X)$ igen får en fordeling med tæthed. Det præcise resultat, der bekræfter denne formodning, er en version af *integraltransformationssætningen*, som bl.a. involverer det analytiske begreb *partiel differentiation* og det algebraiske begreb *determinant*. Vi skriver sætningen op og forklarer den bagefter:

SÆTNING 6.2.2. *Lad E og F være områder i $\mathbf{R}^n$, og lad $t\colon E \to F$ være en bijektiv, kontinuert differentiabel transformation med den egenskab, at $n \times n$–matricen $Dt(x) = \left(\left(\frac{\partial y_i}{\partial x_j}\right)\right)$ af partielle afledede (funktionalmatricen) er regulær (invertibel) for alle x. Definer $q\colon F \to \mathbf{R}$ ved*

$$q(y_1,\ldots,y_n) = \frac{p\left(t^{-1}(y_1,\ldots,y_n)\right)}{|\det Dt\left(t^{-1}(y_1,\ldots,y_n)\right)|}.$$

Da gælder, for enhver pæn delmængde B af F,

$$\begin{aligned}&\int\cdots\int 1_B(y_1,\ldots,y_n)q(y_1,\ldots,y_n)\,dy_1\ldots dy_n\\ &= \int\cdots\int 1_{t^{-1}(B)}(x_1,\ldots,x_n)p(x_1,\ldots,x_n)\,dx_1\ldots dx_n.\end{aligned}$$

Sætningen berettiger således, at vi under disse omstændigheder fortolker $Y = t(X)$ som en n–dimensional kontinuert fordelt variabel med den angivne tæthed q.

Et eksakt bevis for sætningen er teknisk ret krævende, og vil i hvert fald ikke blive givet her. Vi vil indskrænke os til at gennemgå et heuristisk bevis i tilfældet $n = 2$, samtidig med at vi forklarer betydningen af sætningens antagelser. Efter gennemgangen skulle sætningen være til at læse og forstå (hvad den ikke formodes at være i øjeblikket).

Vi lader altså nu E og F betegne områder i planen. Transformationen $t\colon E \to F$ kan skrives på formen

$$\begin{aligned}y_1 &= t_1(x_1,x_2),\\ y_2 &= t_2(x_1,x_2).\end{aligned}$$

Antagelsen om, at t er kontinuert differentiabel, skal forstås på den måde, at enhver af funktionerne t_1 og t_2 er differentiabel som funktion af ethvert af sine to argumenter, når det andet argument holdes fast; således at f.eks. $t_1(x_1,x_2)$ for fastholdt x_2 er en kontinuert differentiabel funktion af x_1. Differentialkvotienten betegnes $\frac{\partial y_1}{\partial x_1}$ og kaldes en *partiel afledt.* Vi vil her anvende betegnelserne

$$a_{ij}(x_1,x_2) = \frac{\partial y_i}{\partial x_j}$$

for disse fire funktioner, således at for eksempel $a_{12}(x_1,x_2)$ er differentialkvotienten af $t_1(x_1,x_2)$ m.h.t. x_2 for fastholdt x_1. Antagelsen om kontinuert differentiabilitet indebærer, ud over hvad der allerede er sagt, at disse fire funktioner $a_{ij}\colon E \to \mathbf{R}$ er kontinuerte funktioner i to variable (og altså ikke kun i den ene variabel der blev differentieret efter, for fastholdt værdi af den anden).

Funktionalmatricen eller *Jacobi matricen* (eller, i mere teknisk jargon, simpelthen *differentialkvotienten*) for t er 2×2–matricen

$$Dt(x) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11}(x_1, x_2) & a_{12}(x_1, x_2) \\ a_{21}(x_1, x_2) & a_{22}(x_1, x_2) \end{bmatrix}.$$

Denne matrix spiller samme rolle som differentialkvotienten i én dimension, forstået på følgende måde: At en sædvanlig reel funktion $y = f(x)$ er differentiabel i punktet x_0 betyder jo, at vi har approksimationen

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

for x nær ved x_0. Den lineære – eller, mere præcist, lineært affine – funktion på højre side svarer til tangenten i punktet $(x_0, f(x_0))$. Tilsvarende gælder i to dimensioner, at den kontinuert differentiable transformation t nær et punkt $x_0 = (x_1^o, x_2^o)$ har den approksimative fremstilling

$$\begin{aligned} t_1(x) &\approx t_1(x_0) + a_{11}(x_0)(x_1 - x_1^o) + a_{12}(x_0)(x_2 - x_2^o), \\ t_2(x) &\approx t_2(x_0) + a_{21}(x_0)(x_1 - x_1^o) + a_{22}(x_0)(x_2 - x_2^o). \end{aligned}$$

Disse approksimative relationer kan sammenfattes på matrixform således:

$$t(x) \approx t(x_0) + Dt(x_0)(x - x_0)$$

(idet det her, som altid når vi regner med matricer, er underforstået at punkter i $\mathbf{R}^n$ (her $\mathbf{R}^2$) opfattes som søjlevektorer). Højre side er en lineært affin funktion af x (dvs. en funktion af typen "konstantled + lineært udtryk"), og betydningen af funktionalmatricen ligger altså i, at den bestemmer den lineære del af denne approksimation.

Lad nu $A = [x_1^o, x_1^o + h_1] \times [x_2^o, x_2^o + h_2]$ være et lille rektangel med $x_0 = (x_1^o, x_2^o)$ som sit nederste venstre hjørne. Billedet $t(A)$ af dette rektangel vil, på nær en deformation som er uden betydning når sidelængderne h_1 og h_2 er små, være et lille parallellogram, hvis ene hjørne er billedpunktet $t(x_0)$. Placeringen af de tre andre hjørner følger af ovenstående approksimationsformler (idet vi fra nu af skriver a_{ij} i stedet for $a_{ij}(x_0)$, for at lette overskueligheden):

$$\begin{bmatrix} t_1(x_1^o + h_1, x_2^o) \\ t_2(x_1^o + h_1, x_2^o) \end{bmatrix} \approx \begin{bmatrix} t_1(x_1^o, x_2^o) + a_{11}h_1 \\ t_2(x_1^o, x_2^o) + a_{21}h_1 \end{bmatrix},$$

$$\begin{bmatrix} t_1(x_1^o, x_2^o + h_2) \\ t_2(x_1^o, x_2^o + h_2) \end{bmatrix} \approx \begin{bmatrix} t_1(x_1^o, x_2^o) + a_{12}h_2 \\ t_2(x_1^o, x_2^o) + a_{22}h_2 \end{bmatrix},$$

$$\begin{bmatrix} t_1(x_1^o + h_1, x_2^o + h_2) \\ t_2(x_1^o + h_1, x_2^o + h_2) \end{bmatrix} \approx \begin{bmatrix} t_1(x_1^o, x_2^o) + a_{11}h_1 + a_{12}h_2 \\ t_2(x_1^o, x_2^o) + a_{21}h_1 + a_{22}h_2 \end{bmatrix}.$$

Anderledes sagt, parallellogrammet $t(A)$ udspændes af vektorerne

$$\begin{bmatrix} a_{11}h_1 \\ a_{21}h_1 \end{bmatrix} \text{ og } \begin{bmatrix} a_{12}h_2 \\ a_{22}h_2 \end{bmatrix}$$

udgående fra punktet $t(x_0)$. Det er en elementær øvelse i geometri (som vi ikke vil gennemføre her) at vise, at arealet af parallellogrammet udspændt af to vektorer er lig med den numeriske værdi af determinanten af den 2×2–matrix der har vektorerne som søjler. Arealet af $t(A)$ bliver således

$$\begin{aligned} |(a_{11}h_1)(a_{22}h_2) - (a_{21}h_1)(a_{12}h_2)| &= h_1h_2|a_{11}a_{22} - a_{21}a_{12}| \\ &= h_1h_2|\det Dt(x_0)|. \end{aligned}$$

Sætning 6.2.2 følger nu af tæthedens fortolkning som det lokale forhold mellem sandsynlighedsmasse og areal. Sandsynligheden for, at en stokastisk variabel $X = (X_1, X_2)$ med tæthed p havner i det lille rektangel A er jo $h_1h_2p(x_0)$. For den afledte variabel $Y = (Y_1, Y_2) = t(X)$ gælder derfor, da t pr. antagelse er bijektiv,

$$P(Y \in t(A)) = P(X \in A) \approx h_1h_2p(x_0),$$

og da arealet af $t(A)$ er $h_1h_2|\det Dt(x_0)|$ må tætheden for Y's fordeling i punktet $t(x_0)$ være

$$q(t(x_0)) = \frac{h_1h_2p(x_0)}{h_1h_2|\det Dt(x_0)|} = \frac{p(x_0)}{|\det Dt(x_0)|},$$

hvilket netop er sætningens påstand.

Et heuristisk bevis for vilkårlig dimension n kan føres på samme måde. Der får man brug for det resultat, at det n–dimensionale volumenmål af et "parallelogram" udspændt af søjlerne i en $n \times n$–matrix er lig med den numeriske værdi af matricens determinant.

Der er en detalje i sætningen vi endnu ikke har forklaret, nemlig antagelsen om at funktionalmatricen skal være regulær i ethvert punkt. Det betyder, at matricen $Dt(x)$ skal have lineært uafhængige søjler, eller (ækvivalent hermed) at den skal svare til en bijektiv lineær afbildning. I dimension 2 betyder dette, at de to vektorer som udspænder det lille parallellogram ikke må være proportionale. Hvis de er det, degenererer parallellogrammet jo til et liniestykke med areal 0. I praksis vil dette undertiden indtræffe f.eks. i punkter på randen af området E. Normalt har det ikke anden betydning, end at tætheden q nærmer sig $+\infty$ i de tilsvarende randpunkter for F. Sætning 6.2.2 gælder i det store og hele alligevel.

Sætningen kan også udvides til tilfældet hvor t ikke er bijektiv. I så fald skal $q(y)$ beregnes ved summation af bidrag fra de punkter x der afbildes over i y, ligesom i dimension 1 (se sidste afsnit af §5.4).

EKSEMPEL 6.2.4 (transformation med lineær afbildning). Lad $t\colon \mathbf{R}^n \to \mathbf{R}^n$ være den lineære afbildning givet ved $n \times n$–matricen T , altså $t(x) = Tx$. En sådan afbildning er naturligvis kontinuert differentiabel med funktionalmatrix $Dt(x) = T$ i alle punkter, og den er bijektiv hvis og kun hvis T er regulær. For en stokastisk variabel X på $\mathbf{R}^n$ med tæthed p får vi, at den afledte variabel $Y = t(X)$ har tætheden

$$q(y) = \frac{p\left(t^{-1}(y)\right)}{|\det T|} = \frac{p\left(T^{-1}y\right)}{|\det T|}.$$

Lidt mere generelt gælder, for en lineært affin afbildning af formen $t(x) = a+Tx$ (a en $n\times 1$ søjlevektor, T en regulær $n\times n$–matrix) at $Y = a+TX$ har tætheden

$$q(y) = \frac{p\left(T^{-1}(y-a)\right)}{|\det T|}.$$

Bemærk at denne formel for $n = 1$ reducerer til formlen for tætheden i en fordeling givet ved positions– og skalaparameter, jvf. eksempel 5.4.3.

Det er lidt bagvendt, at vi således udleder transformationssætningen i det lineært–affine tilfælde som specialtilfælde af den generelle sætning for differentiable transformationer. Som man vil se ved nærlæsning bygger det heuristiske bevis i virkeligheden på sætningens gyldighed i dette specialtilfælde, og i en lidt grundigere fremstilling ville man nok begynde med at bevise sætningen i det lineært–affine tilfælde.

EKSEMPEL 6.2.5. Betragt området

$$E =]0,+\infty[\times]0,2\pi[\subseteq \mathbf{R}^2.$$

På denne "halvstrimmel" kan vi definere en sandsynlighedstæthed ved

$$p(y,v) = \frac{1}{2\pi}e^{-y}.$$

En stokastisk variabel (Y,V) med denne tæthed kan opfattes som et par af uafhængige éndimensionale variable, hvor Y er normeret eksponentialfordelt (faktoren e^{-y}) og V er ligefordelt på intervallet $]0,2\pi[$ (faktoren $\frac{1}{2\pi}$). Definer nu $t\colon E \to \mathbf{R}^2$ ved

$$t(y,v) = \left(\sqrt{2y}\cos(v), \sqrt{2y}\sin(v)\right).$$

Denne transformation er lettest at forstå geometrisk, idet punktet $t(y,v)$ er bestemt ved sine polære koordinater $\sqrt{2y}$ = afstanden fra $(0,0)$ og $v =$

retningsvinklen i forhold til første akse. Heraf følger let at afbildningen t er injektiv. Den er også "næsten surjektiv", idet alle punkter i planen, bortset fra dem med retningsvinkel 0 (den højre halvdel af første akse), er med i billedmængden. Dette lille problem kan man løse ved at opfatte t som afbildning ind i området $F = \mathbf{R}^2 \setminus \{(x_1, 0) \mid x_1 \geq 0\}$. Men problemer vedrørende sådanne singularitetsmængder, som alligevel er uden betydning da de rammes med sandsynlighed 0 af en kontinuert fordelt variabel, kan normalt ignoreres, og det vil vi gøre i det følgende.

Transformationen t er naturligvis kontinuert differentiabel, og funktionalmatricen ser sådan ud:

$$Dt(y,v) = \begin{bmatrix} \frac{1}{\sqrt{2y}}\cos(v) & -\sqrt{2y}\sin(v) \\ \frac{1}{\sqrt{2y}}\sin(v) & \sqrt{2y}\cos(v) \end{bmatrix}.$$

Funktionaldeterminanten (= funktionalmatricens determinant) er således

$$\left(\frac{1}{\sqrt{2y}}\cos(v)\right)\left(\sqrt{2y}\cos(v)\right) + \left(\frac{1}{\sqrt{2y}}\sin(v)\right)\left(\sqrt{2y}\sin(v)\right) = 1$$

(hvilket ikke er noget tilfælde, det var faktisk planlagt). Af sætning 6.2.2 følger så, at den transformerede stokastiske variable

$$(X_1, X_2) = t(Y, V) = \left(\sqrt{2Y}\cos(V), \sqrt{2Y}\sin(V)\right)$$

har tætheden

$$q(x_1, x_2) = p\left(t^{-1}(x_1, x_2)\right) = \frac{1}{2\pi}e^{-(x_1^2+x_2^2)/2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}e^{-x_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-x_2^2/2}\right).$$

De to variable X_1 og X_2 er altså stokastisk uafhængige, normeret normalfordelte.

Hermed har vi ud fra en rektangulært fordelt variabel og en eksponentialfordelt variabel konstrueret to normalfordelte variable. Det er et trick, man kan benytte til simulation af normalfordelte variable. I princippet kan normalfordelte variable $X_1, X_2, \ldots$ konstrueres ud fra almindelige tilfældige tal (dvs. rektangulært fordelte variable på enhedsintervallet) efter skemaet $X_1 = \Phi^{-1}(R_1)$, $X_2 = \Phi^{-1}(R_2), \ldots$. Men det kræver, at man har en procedure til beregning af Φ^{-1}, og selv om man måske nok kan få fat i sådan en, vil den med garanti være meget langsommere end de procedurer man har til udregning af logaritmefunktion og trigonometriske funktioner. Så det betaler sig i stedet at benytte skemaet

$$\begin{aligned}
X_1 &= \sqrt{-2\log(R_1)}\cos(2\pi R_2) \\
X_2 &= \sqrt{-2\log(R_1)}\sin(2\pi R_2) \\
X_3 &= \sqrt{-2\log(R_3)}\cos(2\pi R_4) \\
X_4 &= \sqrt{-2\log(R_3)}\sin(2\pi R_4) \\
&\vdots
\end{aligned}$$

Man kan eventuelt nøjes med at benytte hveranden af disse variable, således at hver normalfordelt variabel genereres ved hjælp af to tilfældige tal; i pascal f.eks. ved et direktiv af typen

```
NormFordVar := sqrt(-2*ln(random))*sin(2*pi*random);
```

Eftersom transformationen t er bijektiv, kan det vi har vist også opfattes som et resultat vedrørende t^{-1}. Så kommer det til at se sådan ud:

Lad X_1 og X_2 være uafhængige, normeret normalfordelte variable, og definer

$$\begin{aligned}
Y &= \frac{1}{2}\left(X_1^2 + X_2^2\right), \\
V &= \arg\left(X_1, X_2\right),
\end{aligned}$$

hvor arg *her betegner den funktion, der til en vektor (x_1, x_2) knytter dens retningsvinkel $v \in [0, 2\pi[$ med x_1–aksen. Så er Y og V stokastisk uafhængige, Y normeret eksponentialfordelt og V ligefordelt på intervallet $[0, 2\pi[$.*

Det mest interessante ved dette eksempel er måske, at vi nu endelig har indset at normeringsfaktoren $\frac{1}{\sqrt{2\pi}}$ i den normale fordelings tæthed er korrekt. I §5.6 måtte vi nøjes med at postulere det. Men havde vi der valgt blot at give den et navn, f.eks. c, ville vi efter gennemregning af dette eksempel let kunne slutte at der gælder $c^2 = \frac{1}{2\pi}$.

Vi slutter denne paragraf med et eksempel som viser, at der findes andre metoder til udledning af transformerede fordelinger end dem vi har nævnt.

Eksempel 6.2.6. Lad $X_1, \ldots, X_n$ være stokastisk uafhængige, rektangulært fordelte på $[0, 1]$. Med $X_{(1)}$ betegnes den mindste af disse, altså

$$X_{(1)} = X_1 \wedge X_2 \wedge \cdots \wedge X_n.$$

I dette tilfælde kan vi direkte opskrive fordelingsfunktionen for $X_{(1)}$, idet der jo gælder

$$\begin{aligned} P\left(X_{(1)} \leq x\right) &= 1 - P\left(X_{(1)} > x\right) \\ &= 1 - P\left(\text{alle variable } X_1, \dots, X_n \text{ er } > x\right) \\ &= 1 - P\left(X_1 > x\right) \dots P\left(X_n > x\right) \\ &= 1 - (1-x)^n. \end{aligned}$$

Ved differentiation fås tætheden for $X_{(1)}$ til

$$q(x) = n(1-x)^{n-1} \qquad (0 \leq x \leq 1).$$

Mere generelt kunne vi interessere os for fordelingen af én af de ordnede variable $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, som fremkommer ved at de oprindelige variable $X_1, \dots, X_n$ ordnes efter størrelse. For den "k–te mindste" $X_{(k)}$ ses at

$$\begin{aligned} &P\left(X_{(k)} \leq x\right) \\ &= P\left(\text{mindst } k \text{ af de } n \text{ variable er } \leq x\right) \\ &= P\left(\#\{i \mid X_i \leq x\} \geq k\right). \end{aligned}$$

For hver af de oprindelige variable gælder her at $P\left(X_i \leq x\right) = x$, og da de er uafhængige, vil antallet iblandt dem som er $\leq x$ åbenbart være binomialfordelt med sandsynlighedsparameter x og antalsparameter n. Heraf fås følgende udtryk for $X_{(k)}$'s fordelingsfunktion:

$$P\left(X_{(k)} \leq x\right) = \sum_{s=k}^{n} \binom{n}{s} x^s (1-x)^{n-s}.$$

En anden mulighed består i direkte at udregne fordelingens tæthed efter følgende argument. For et lille interval $[x, x+h[$, betragt antallene

$$\begin{aligned} S_< &= \#\{i \mid X_i < x\}, \\ S_h &= \#\{i \mid x \leq X_i < x+h\}, \\ S_> &= \#\{i \mid x+h \leq X_i\}. \end{aligned}$$

Da er $(S_<, S_h, S_>)$ åbenbart polynomialfordelt med antalsparameter n og sandsynlighedsparametre $p_< = x$, $p_h = h$ og $p_> = 1 - x - h$. Heraf følger (for $h \to 0$, idet vi ignorerer den hændelse at to eller flere af de variable X_i falder i det lille interval) at

$$\begin{aligned} P\left(X_{(k)} \in [x, x+h[\right) &\approx P\left(S_< = k-1, S_h = 1, S_> = n-k\right) \\ &= \begin{pmatrix} & n & \\ k-1 & 1 & n-k \end{pmatrix} x^{k-1} h (1-x-h)^{n-k} \\ &\approx h \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k}, \end{aligned}$$

dvs. tætheden for $X_{(k)}$ er (ifølge den sædvanlige “lokale” fortolkning)

$$q(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}.$$

Det er ikke umiddelbart indlysende, at den fundne fordelingsfunktion ved differentiation giver denne tæthed, men det gør den altså.

Opgave 6.2.1. Lad X_1 og X_2 være stokastisk uafhængige, normeret normalfordelte. Vis at fordelingen af X_1/X_2 er den normerede Cauchy fordeling (jvf. opgave 5.4.4).

Opgave 6.2.2*. (Poissonprocessen). Vi har tidligere, i forbindelse med vores cyklist C og tre opgaver vedrørende taxaankomster (eksempel 3.6.2 og 3.6.3, opgave 3.6.1, 3.7.3 og 5.4.3) betragtet situationer, der generelt kan beskrives således: Til ethvert interval $[s, t]$ på tidsskalaen $[0, +\infty[$ er knyttet en stokastisk variabel $N(s, t)$ på $\mathbf{N}_0$, der fortolkes som antallet af hændelser (punkteringer, taxaankomster) som finder sted i det pågældende tidsinterval. Den stokastiske variabel $N(s, t)$ antages Poissonfordelt med parameter $\lambda \times (t - s)$, hvor λ (= det forventede antal hændelser pr. tidsenhed) kaldes *intensiteten*. Endvidere antages, at antal hændelser hørende til disjunkte tidsintervaller er stokastisk uafhængige. I det følgende antages for nemheds skyld at $\lambda = 1$ (hvilket blot betyder, at vi benytter en tidsenhed, der svarer naturligt til situationen).

Vi har tidligere (eksempel 5.1.2) konstateret, at ventetiden T_1 til første hændelse er eksponentialfordelt (i dette tilfælde normeret, da $\lambda = 1$).

(a) Vis at ventetiden T_n til n'te hændelse indtræffer, altså

$$T_n = \inf\{t \mid N(0, t) \geq n\},$$

følger en fordeling med tæthed

$$\frac{1}{(n-1)!} t^{n-1} e^{-t}.$$

(Vink: Udtryk hændelsen $T_n \in [t, t + h[$ ved $N(0, t)$ og $N(t, t + h)$, og beregn dens sandsynlighed, idet muligheden for to eller flere hændelser i det lille interval som sædvanligt kan ignoreres. Sammenlign i øvrigt med resultatet af opgave 3.7.3).

(b) Vis, ved et tilsvarende argument i n dimensioner, at de succesive ventetider

$$\begin{aligned} S_1 &= T_1, \\ S_2 &= T_2 - T_1, \\ &\vdots \\ S_n &= T_n - T_{n-1} \end{aligned}$$

(altså for eksempel tidsrummene mellem taxaankomster) er stokastisk uafhængige, identisk fordelte. (Vink: Udled først fordelingen af (T_1, ..., T_n) ; metoden fra spørgsmål (a) kan benyttes omtrent uændret, idet man for at finde den n–dimensionale fordelings tæthed udregner sandsynligheden for hændelsen

$$\{t_1 \leq T_1 \leq t_1 + h\} \cap \cdots \cap \{t_n \leq T_n \leq t_n + h\}$$

$$= \{N(0, t_1) = 0\} \cap \{N(t_1, t_1 + h) = 1\} \cap \ldots \\ \cap \{N(t_{n-1} + h, t_n) = 0\} \cap \{N(t_n, t_n + h) = 1\}.$$

Udled derefter fordelingen af $(S_1, \ldots, S_n)$ ved transformation).

Opgave 6.2.3. (Illustration af argumentationen i eksempel 6.2.6). Lad $X_1, \ldots, X_5$ være uafhængige, rektangulært fordelte på $[0, 1]$.

(a) Hvad er fordelingen af (S_1, S_2, S_3), hvor

$$\begin{aligned} S_1 &= \#\{i | X_i \in [0, x[\} \\ S_2 &= \#\{i | X_i \in [x, x + h[\} \\ S_3 &= \#\{i | X_i \in [x + h, 1]\} \end{aligned}$$

(b) Hvad er $P((S_1, S_2, S_3) = (2, 1, 2))$?

(c) Hvad er tætheden for fordelingen af $X_{(3)}$ (= det midterste af de fem tal $X_1, \ldots, X_5$)? (Vink: Sandsynligheden for $X_{(3)} \in [x, x + h]$ er, approksimativt for h lille, den vi lige har regnet ud).

Opgave 6.2.4*. Lad X_1, ..., X_n være uafhængige, identisk fordelte med en tæthed p som er strengt positiv på hele aksen (eller et interval). Lad F betegne den tilhørende fordelingsfunktion. Med $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ betegnes de ordnede variable, jvf. eksempel 6.2.6. Vis, at fordelingen af $X_{(k)}$ har tætheden

$$q(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} (1 - F(x))^{n-k} p(x).$$

(Vink: De tranformerede variable $R_i = F(X_i)$ er ligefordelte på enhedsintervallet, og fordelingen af $R_{(k)} = F(X_{(k)})$ er derfor den der blev udledt i eksempel 6.2.6).

Opgave 6.2.5. Lad X_1 og X_2 være stokastisk uafhængige, normeret eksponentialfordelte. Vis at de to afledte variable

$$Y = \frac{X_1}{X_2}$$

$$S = X_1 + X_2$$

er stokastisk uafhængige, og angiv deres fordelinger.

6.3. Foldning.

Vi har ikke sagt meget om, hvordan man håndterer dimensionsreducerende transformationer $t\colon \mathbf{R}^n \to \mathbf{R}^k$, $k < n$, når det ikke netop drejer sig om en projektion ned på nogle af koordinatakserne. Men en simpel metode går ud på følgende: Vælg – med større eller mindre snedighed – en "supplerende" transformation $s\colon \mathbf{R}^n \to \mathbf{R}^{n-k}$, således at

$$(t,s)\colon \mathbf{R}^n \to \mathbf{R}^k \times \mathbf{R}^{n-k} = \mathbf{R}^n$$

opfylder sætning 6.2.2's betingelser. Herefter er man, i hvert fald i princippet, i stand til at opskrive tætheden for fordelingen af $(t(X), s(X))$, og ud fra denne kan tætheden for $t(X)$ ifølge sætning 6.2.1 udregnes, ved at man integrerer de sidste $n-k$ variable ud. Den simplest mulige anvendelse af denne teknik indgår i beviset for følgende resultat:

Sætning 6.3.1. *Lad X_1 og X_2 være reelle stokastiske variable, uafhængige med tætheder p_1 og p_2. Så har fordelingen af $Y = X_1 + X_2$ tætheden*

$$q(y) = \int p_1(y-x)p_2(x)\,dx.$$

Bemærk analogien med formlen i opgave 3.5.4, som handler om det tilsvarende resultat for fordelinger på $\mathbf{N}_0$.

Bevis. Vi deler transformationen $(x_1, x_2) \to y$ op i to efter skemaet

$$(x_1, x_2) \longrightarrow (y, x_2) \longrightarrow y.$$

Den første af disse er den lineære transformation $\mathbf{R}^2 \to \mathbf{R}^2$ givet ved

$$\begin{aligned} y &= x_1 + x_2 \\ x_2 &= \phantom{x_1 + {}} x_2 \end{aligned}$$

hvis tilhørende matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ har determinant 1. Heraf følger (jvf. eksempel 6.2.4) at fordelingen af (Y, X_2) har tætheden

$$r(y, x_2) = p_1(y - x_2)p_2(x_2).$$

Ifølge sætning 6.2.1 fås tætheden for fordelingen af y ved at integrere dette udtryk m.h.t. x_2, hvilket umiddelbart giver den ønskede påstand.

Fordelingen af en sum af to uafhængige variable kaldes *foldningen* af de to variables fordelinger. Integraler af den type, der indgår i sætningens påstand, kaldes *foldningsintegraler.*

Sætning 6.3.2 *(Den normale fordelings foldningsegenskab). Lad X_1, $\ldots$, X_n være uafhængige, normalfordelte med middelværdier $\mu_1, \ldots, \mu_n$*

og varianser $\sigma_1^2, \ldots, \sigma_n^2$. Da er $X_1 + \cdots + X_n$ igen normalfordelt med middelværdi $\mu_1 + \cdots + \mu_n$ og varians $\sigma_1^2 + \cdots + \sigma_n^2$.

Bemærkninger. Det er ikke nogen overraskelse, at middelværdi og varians for summen bliver som angivet, for det er en konsekvens af generelle regneregler for middelværdier og varianser (sætning 4.2.2 og sætning 4.4.2). Det interessante er, at fordelingen af summen igen er normal. Dette resultat har nøje sammenhæng med den centrale grænseværdisætning. Faktisk følger det af den centrale grænseværdisætning, løst sagt fordi en sum af summer af mange små uafhængige bidrag igen kan opfattes som en sum af mange små uafhængige bidrag. Men dette er at vende op og ned på tingene, for ethvert bevis for den centrale grænseværdisætning bygger (mere eller mindre åbenlyst) på den normale fordelings foldningsegenskab. Det er mere korrekt at sige, at det er den normale fordelings foldningsegenskab som gør, at netop denne fordeling optræder som grænsefordeling for summer af uafhængige variable. Man kan vise, at den normale fordeling er den eneste fordeling med endelig varians, som har den egenskab, at foldning af fordelinger af dens type igen fører til en fordeling af denne type.

BEVIS. Bemærk først, at det er nok at se på tilfældet $n = 2$, da det generelle resultat umiddelbart følger ved induktion.

Man kan vise sætningen ved direkte udregning af foldningsintegralet. Men algebraen i forbindelse med omskrivningen af andengrads polynomiet i eksponenten er lidt besværlig. En lidt mere elegant metode er følgende:

Vi begynder med at bemærke, at hvis U_1 og U_2 er uafhængige, normeret normalfordelte, så er, for ethvert talpar (α, β) med $\alpha^2 + \beta^2 = 1$, linearkombinationen $U = \alpha U_1 + \beta U_2$ igen normeret normalfordelt. Dette følger af, at vi kan supplere transformationen $(u_1, u_2) \to u = \alpha u_1 + \beta u_2$ med en anden af samme type til den lineære afbildning $(u_1, u_2) \to (u, v)$ givet ved

$$\begin{aligned} u &= \alpha u_1 + \beta u_2 \\ v &= -\beta u_1 + \alpha u_2 \end{aligned}$$

som ses at være en rotation. Det er intuitivt indlysende, at fordelingen af (U_1, U_2) er invariant under rotationer (tætheden er jo en funktion af afstanden fra $(0, 0)$), og det følger også let af transformationssætningen for det lineære tilfælde (eksempel 6.2.4).

Lad nu X_1 og X_2 være uafhængige, normalfordelte med middelværdier μ_1 og μ_2 og varianser σ_1^2 og σ_2^2. Betragt de tilsvarende standardiserede variable

$$U_1 = \frac{X_1 - \mu_1}{\sigma_1} \qquad \text{og} \qquad U_2 = \frac{X_2 - \mu_2}{\sigma_2},$$

som er uafhængige og normeret normalfordelte. Omskrivningen

$$\begin{aligned} X_1 + X_2 &= (\mu_1 + \sigma_1 U_1) + (\mu_2 + \sigma_2 U_2) \\ &= (\mu_1 + \mu_2) + (\sigma_1 U_1 + \sigma_2 U_2) \\ &= (\mu_1 + \mu_2) + \sqrt{\sigma_1^2 + \sigma_2^2} \left(\frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} U_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} U_2 \right) \end{aligned}$$

viser, idet kvadratsummen af koefficienterne i sidste parentes netop er 1, at $X_1 + X_2$ har formen

$$X_1 + X_2 = \mu_1 + \mu_2 + \sqrt{\sigma_1^2 + \sigma_2^2}\, U,$$

hvor U er normeret normalfordelt. Heraf følger sætningen umiddelbart.

EKSEMPEL 6.3.1 *(Cauchyfordelingens foldningsegenskab).* Som et rent kuriosum vil vi – uden antydning af bevis (det er nemlig svært) – nævne følgende ejendommmelige egenskab ved Cauchy fordelingen (opgave 5.4.4): Hvis $X_1, \ldots, X_n$ er uafhængige, normeret Cauchyfordelte, så er $\frac{1}{n}(X_1 + \cdots + X_n)$ igen normeret Cauchyfordelt (bemærk: i den tilsvarende situation for normalfordelingen er det $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ som er normeret normalfordelt). Vi har her et ekstremt eksempel på, at de store tals lov ikke gælder i nogen som helst forstand for variable, der ikke har middelværdi. Hvis man forestillede sig observationer af en ukendt størrelse θ, behæftede med uafhængige Cauchyfordelte fejl, således at de størrelser man i virkeligheden observerede var $\theta + X_1$, $\theta + X_2$, ..., så ville man være ude af stand til at forbedre sit skøn over θ ved at tage gennemsnit af mange observationer, fordi ethvert sådant gennemsnit ville have samme fordeling som en enkelt observation.

OPGAVE 6.3.1. Lad X_1, X_2 og X_3 være uafhængige, rektangulært fordelte på enhedsintervallet.

(a) Udregn og tegn tætheden for fordelingen af $X_1 + X_2$. Bemærk sammenhængen med eksempel 6.2.2, og bemærk også analogien med det tilsvarende diskrete problem, som vi kender fra eksempel 1.2.1.

(b) Udregn og tegn tætheden for fordelingen af $X_1 + X_2 + X_3$.

(c) Udregn og tegn tætheden for fordelingen af $X_1 + 10X_2$.

OPGAVE 6.3.2. Lad X_1 og X_2 være uafhængige, normeret eksponentialfordelte. Udregn tætheden for fordelingen af $X_1 - X_2$.

OPGAVE 6.3.3. Opgave 6.2.2 indeholder, som biprodukt, et resultat vedrørende foldning af normerede eksponentialfordelinger. Formuler dette, og bevis det direkte ved induktion. Sammenlign med opgave 3.7.2, som indeholder det tilsvarende resultat i det diskrete tilfælde (den geometriske fordeling).

OPGAVE 6.3.4. Lad X_1 og X_2 være uafhængige, normeret eksponentialfordelte. Definer

$$M = \min\{X_1, X_2\} = X_1 \wedge X_2,$$
$$S = \max\{X_1, X_2\} = X_1 \vee X_2,$$
$$D = S - M = |X_1 - X_2|.$$

Angiv (ved tætheder)

(a) Fordelingen af M.

(b) Fordelingen af (M, S).

(c) Fordelingen af (M, D).

(d) Fordelingen af S.

OPGAVE 6.3.5. Lad R_1 og R_2 være uafhængige, rektangulært fordelte på enhedsintervallet. Sæt $Z = R_1 R_2$.

(a) Hvad er fordelingen af (R_1, Z)?

(b) Hvad er fordelingen af Z?

Fordelingen af Z kan man også nå frem til på følgende måde:

(c) Sæt $X_1 = -\log(R_1)$ og $X_2 = -\log(R_2)$. Hvad er fordelingen af (X_1, X_2)?

(d) Udled fordelingen af $S = X_1 + X_2$.

(e) Udled fordelingen af $Z = \exp(-S)$.

OPGAVE 6.3.6. Lad R og X være uafhængige, R rektangulært fordelt på enhedsintervallet, X med en vilkårlig tæthed p. Lad F betegne fordelingsfunktionen for X. Vis at tætheden for $S = R + X$ kan skrives

$$q(s) = F(s) - F(s-1).$$

Udregn og tegn tætheden q i tilfældet hvor X er normeret eksponentialfordelt.