Kapitel 7

Den normale fordelings teori

7.1. Gamma–fordelingen og beta–fordelingen.

Vi begynder med at indføre nogle fordelinger, som tilsyneladende ikke har meget med normalfordelingen at gøre. Men det har de faktisk, som vi senere skal se.

Γ*–fordelingen.* Betragt integralet

$$\Gamma(\lambda) = \int_0^{+\infty} x^{\lambda-1} e^{-x}\, dx,$$

som let ses at være veldefineret for $\lambda > 0$. Som funktion af λ kaldes dette integral Γ*–funktionen.* Normering af integranden ved division med $\Gamma(\lambda)$ fører til en klasse af tætheder, hvis tilhørende fordelinger kaldes Γ*–fordelinger.* Parameteren λ kaldes her *formparameteren* (idet den i modsætning til eventuelle skala– og positionsparametre har indflydelse på tæthedens form). Γ–fordelingen med formparameter λ er således givet ved tætheden

$$p(x) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x} \quad (x \geq 0).$$

Vi har allerede mødt disse fordelinger for heltallige værdier af λ i opgave 6.2.2 (a), hvor Γ–fordelingen med formparameter n optrådte som fordelingen af ventetiden til n'te hændelse i en Poisson proces. Dér så normeringsfaktoren lidt anderledes ud, men eftersom der kun er én måde, hvorpå en funktion kan normeres til en tæthed, må vi konkludere, at der for $\lambda \in \mathbf{N}$ gælder

$$\Gamma(\lambda) = (\lambda - 1)!.$$

Dette følger også af relationen

$$\Gamma(\lambda + 1) = \lambda \Gamma(\lambda),$$

som kaldes Γ*–funktionens funktionalligning,* i forbindelse med den trivielle observation at $\Gamma(1) = 1$. Γ–funktionens funktionalligning er let at vise ved delvis integration:

$$\begin{aligned}
\Gamma(\lambda + 1) &= \int_0^{+\infty} x^{\lambda} e^{-x}\, dx \\
&= \left[x^{\lambda}\left(-e^{-x}\right)\right]_0^{+\infty} - \int_0^{+\infty} \lambda x^{\lambda-1}(-e^{-x})\, dx \\
&= \lambda \Gamma(\lambda).
\end{aligned}$$

B–*fordelingen* (udtales beta–fordelingen, idet B'et skal forestille et stort græsk bogstav). Integralet

$$\mathrm{B}(\lambda_1, \lambda_2) = \int_0^1 x^{\lambda_1 - 1}(1-x)^{\lambda_2 - 1}\,dx$$

ses at være veldefineret for $\lambda_1, \lambda_2 > 0$. Ved normering fås således en sandsynlighedstæthed

$$p(x) = \frac{1}{\mathrm{B}(\lambda_1, \lambda_2)}\, x^{\lambda_1 - 1}(1-x)^{\lambda_2 - 1}$$

for en fordeling på enhedsintervallet. Denne fordeling kaldes B–*fordelingen* med (form–) parametre λ_1 og λ_2.

Vi har truffet B–fordelingen for heltallige parameterværdier i eksempel 6.2.6, hvor den for $\lambda_1 = k$ og $\lambda_2 = n - k + 1$ optrådte som fordelingen af den k'te i det ordnede sæt, dannet ud fra n ligefordelte variable på enhedsintervallet. Dér havde vi et andet udtryk for normeringsfaktoren, hvilket sætter os i stand til at konkludere at der for $\lambda_1, \lambda_2 \in \mathbf{N}$ gælder

$$\mathrm{B}(\lambda_1, \lambda_2) = \frac{(\lambda_1 - 1)!(\lambda_2 - 1)!}{(\lambda_1 + \lambda_2 - 1)!}.$$

Sætning 7.1.1. *Lad X_1 og X_2 være stokastisk uafhængige, Γ–fordelte med formparametre λ_1 og λ_2. Definer*

$$S = X_1 + X_2,$$
$$Z = \frac{X_1}{X_1 + X_2}.$$

Da er S og Z stokastisk uafhængige, S Γ–fordelt med formparameter $\lambda_1 + \lambda_2$, Z B–fordelt med parametre (λ_1, λ_2).

Bevis. Transformationen

$$\begin{aligned} (x_1, x_2) &\longrightarrow (s, z) = \left(x_1 + x_2, \frac{x_1}{x_1 + x_2}\right) \\ ]0,+\infty[\times]0,+\infty[&\longrightarrow]0,+\infty[\times]0,1[\end{aligned}$$

ses let at være bijektiv. Den er også kontinuert differentiabel, med funktionalmatrix

$$\begin{bmatrix} 1 & 1 \\ \frac{x_2}{(x_1+x_2)^2} & \frac{-x_1}{(x_1+x_2)^2} \end{bmatrix}.$$

Funktionaldeterminantens numeriske værdi er således

$$\left| \frac{-x_1}{(x_1+x_2)^2} - \frac{x_2}{(x_1+x_2)^2} \right| = \frac{1}{x_1 + x_2}.$$

Tætheden for fordelingen af (S, Z) får vi, ifølge sætning 6.2.2, ved at opskrive tætheden for (X_1, X_2) divideret med funktionaldeterminantens numeriske værdi, og omskrive dette udtryk i x_1 og x_2 til et udtryk i de nye variable s og z:

$$\begin{aligned}
&\left(\frac{1}{\Gamma(\lambda_1)} x_1^{\lambda_1-1} e^{-x_1}\right)\left(\frac{1}{\Gamma(\lambda_2)} x_2^{\lambda_2-1} e^{-x_2}\right)(x_1+x_2)\\
&= \frac{1}{\Gamma(\lambda_1)}(sz)^{\lambda_1-1}\frac{1}{\Gamma(\lambda_2)}(s(1-z))^{\lambda_2-1} s e^{-s}\\
&= \frac{1}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\left(s^{\lambda_1+\lambda_2-1}e^{-s}\right)\left(z^{\lambda_1-1}(1-z)^{\lambda_2-1}\right).
\end{aligned}$$

Dette udtryk ser ud som ønsket, bortset fra at normeringsfaktoren skulle have været produkt af $\frac{1}{\Gamma(\lambda_1+\lambda_2)}$ (stammende fra Γ–fordelingen af S) og $\frac{1}{\mathrm{B}(\lambda_1,\lambda_2)}$ (stammende fra B–fordelingen af Z). Men igen, da en funktion kun på én måde kan normeres til en tæthed, har vi ikke alene vist sætningen, men også relationen

$$\frac{1}{\Gamma(\lambda_1)\Gamma(\lambda_2)} = \frac{1}{\Gamma(\lambda_1+\lambda_2)}\frac{1}{\mathrm{B}(\lambda_1,\lambda_2)}$$

eller

$$\mathrm{B}(\lambda_1,\lambda_2) = \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1+\lambda_2)}$$

der udtrykker B–funktionen ved Γ–funktionen (og som vi i øvrigt allerede har stiftet bekendtskab med for heltallige parameterværdier).

Yderligere har vi hermed bevist, at foldning af to Γ–fordelinger fører til en ny Γ–fordeling, med en formparameter der er sum af de to oprindelige formparametre. Dette resultat udvides let ved induktion til følgende:

SÆTNING 7.1.2 *(Γ–fordelingens foldningsegenskab). Lad $X_1, \ldots, X_n$ være uafhængige, Γ–fordelte med formparametre $\lambda_1, \ldots, \lambda_n$. Da er $X_1 + \cdots + X_n$ Γ–fordelt med formparameter $\lambda_1 + \cdots + \lambda_n$.*

Også dette resultat har vi set før i tilfælde af heltallige formparametre (opgave 6.2.2 og 6.3.3).

Γ– og B–fordelingerne spiller en rolle i forbindelse med statistisk analyse af normalfordelte observationer. Følgende to sætninger giver nøglen til denne sammenhæng:

SÆTNING 7.1.3. *Lad $U_1, \ldots, U_n$ være stokastisk uafhængige, normeret normalfordelte. Da er $X = U_1^2 + \cdots + U_n^2$ Γ–fordelt med formparameter $n/2$ og skalaparameter 2.*

I de lineære normalfordelingsmodellers teori er denne fordeling så vigtig, at den har sit eget navn. I statistisk litteratur kaldes den χ^2*–fordelingen* *med n frihedsgrader.* En χ^2–fordeling er altså en Γ–fordeling med skalaparameter 2, og dens formparameter λ angives ved "frihedsgradsantallet" 2λ (som altid er et helt tal).

Bemærk at vi allerede har vist sætningen for $n = 2$ (i eksempel 6.2.5).

BEVIS. På grund af Γ–fordelingens foldningsegenskab er det åbenbart tilstrækkeligt at vise sætningen for $n = 1$. Vi skal altså vise, at hvis U er normeret normalfordelt, så er $X = U^2/2$ Γ–fordelt med $\lambda = \frac{1}{2}$. Dette er en simpel anvendelse af sætning 5.4.1. Man skal blot være opmærksom på, at da transformationen $u \to u^2/2$ ikke er bijektiv, skal tætheden i et punkt x beregnes ved summation af bidrag fra de punkter u der afbildes over i x. I dette tilfælde betyder det blot at man skal gange med faktoren 2, idet både tætheden og den numeriske værdi af differentialkvotienten er den samme i de to punkter $u = \pm\sqrt{2x}$ som afbildes over i et givet x. Tæthedens værdi i punktet $x = u^2/2$ er derfor

$$2\frac{\frac{1}{\sqrt{2\pi}}e^{-u^2/2}}{|u|} = 2\frac{\frac{1}{\sqrt{2\pi}}e^{-x}}{\sqrt{2x}} = \frac{1}{\sqrt{\pi}}x^{\frac{1}{2}-1}e^{-x}.$$

Heraf følger sætningen umiddelbart; desuden har vi — endnu engang — fået en normeringsfaktor, der ikke så ud som forventet, og endnu engang kan dette bruges til at udlede en ny relation. Her følger det jo at

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

Ved hjælp af Γ–funktionens funktionalligning kan man herefter udregne værdierne i alle andre punkter af formen $n + \frac{1}{2}$:

$$\begin{aligned}
\Gamma\left(\frac{3}{2}\right) &= \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}, \\
\Gamma\left(\frac{5}{2}\right) &= \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi}, \\
&\vdots \\
\Gamma\left(n+\frac{1}{2}\right) &= \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\dots\frac{3}{2}\frac{1}{2}\sqrt{\pi}.
\end{aligned}$$

Ved kombination af sætningerne 7.1.1 og 7.1.3 får man følgende resultat, som viser sig at være vigtigt i forbindelse med de statistiske anvendelser:

SÆTNING 7.1.4. *Lad $U_1, \dots, U_n$ være uafhængige, normeret normalfordelte. Betragt for $k < n$ de to afledte variable*

$$Z = \frac{U_1^2 + \cdots + U_k^2}{U_1^2 + \cdots + U_n^2}$$

og

$$S = U_1^2 + \cdots + U_n^2.$$

Da er Z og S stokastisk uafhængige. Z er B*–fordelt med parametre $\frac{k}{2}$ og $\frac{n-k}{2}$, og S er χ^2–fordelt med n frihedsgrader.*

I de statistiske anvendelser plejer man her, i stedet for den B–fordelte variabel Z, at betragte

$$V = \frac{(U_1^2 + \cdots + U_k^2)/k}{(U_{k+1}^2 + \cdots + U_n^2)/(n-k)}.$$

Dette svarer blot til en entydig transformation, idet

$$Z = \frac{kV}{(n-k)+kV} \qquad \text{og} \qquad V = \frac{(n-k)Z}{k(1-Z)}.$$

Fordelingen af V kaldes F–*fordelingen med* $(k, n-k)$ *frihedsgrader.* Bemærk at den, på nær en skalaparameter, kan fortolkes som fordelingen af forholdet mellem to uafhængige χ^2–fordelte variable med henholdsvis k og $n-k$ frihedsgrader.

I tilfældet $k = 1$ betragter man oftere variablen

$$T = \frac{U_1}{\sqrt{(U_2^2 + \cdots + U_n^2)/(n-1)}}$$

hvis fordeling kaldes T–*fordelingen med* $n-1$ *frihedsgrader.* Bemærk at der gælder $T^2 = V$, så bortset fra T's fortegn (som følger fortegnet for U_1) er der igen tale om en transformation af V (eller Z). Når frihedsgradsantallet $n-1$ er stort ligner T–fordelingen en normeret normalfordeling. Intuitivt følger dette af, at nævneren i udtrykket for T ifølge de store tals lov konvergerer mod 1, således at $T \approx U_1$ for n stor.

OPGAVE 7.1.1*. Lad X være Γ–fordelt med formparameter λ. Vis at $\mathrm{E}X = \mathrm{var}(X) = \lambda$. Hvad er middelværdi og varians i χ^2–fordelingen med n frihedsgrader ? Tegn tæthederne for χ^2–fordelinger med 1, 2, 3 og 4 frihedsgrader.

OPGAVE 7.1.2*. Lad Z være B–fordelt med parametre (λ_1, λ_2). Vis at

$$\mathrm{E}Z = \frac{\lambda_1}{\lambda_1 + \lambda_2},$$

$$\mathrm{var}(Z) = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2(\lambda_1 + \lambda_2 + 1)}.$$

Tegn B–fordelingens tæthed for $(\lambda_1, \lambda_2) = (0.5, 0.5)$, $(1, 1)$, $(1.5, 1.5)$, $(2, 2)$, $(3, 3)$ og $(4, 6)$. To af disse fordelinger har vi truffet i anden sammenhæng. Hvor?

OPGAVE 7.1.3*.

(a) Vis at T–fordelingen med f frihedsgrader har tæthed

$$p(t) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\sqrt{\pi f}\,\Gamma\left(\frac{f}{2}\right)\left(1+\frac{t^2}{f}\right)^{\frac{f+1}{2}}}.$$

(Vink: Ifølge sin definition er T–fordelingen fordelingen af

$$T = \frac{U}{\sqrt{\frac{1}{f}\left(U_1^2+\cdots+U_f^2\right)}}$$

hvor $U, U_1, \ldots, U_f$ er uafhængige, normeret normalfordelte. Ifølge sætning 7.1.4 er

$$Z = \frac{U^2}{U^2+(U_1^2+\cdots+U_f^2)} = \frac{T^2}{T^2+f} = \frac{|T|^2}{|T|^2+f}$$

B–fordelt med parametre $(\frac{1}{2}, \frac{f}{2})$. Da $|T|$ afhænger entydigt af Z kan tætheden for $S = |T|$ beregnes, og da fordelingen af T åbenbart er symmetrisk om 0 er det herefter ikke svært at opskrive dennes tæthed. Bemærk, at det til udledning af fordelingen af $S = |T|$ ikke er nødvendigt at skrive $s = |t|$ som funktion af z og differentiere. Det er lettere, da det hele alligevel skal skrives som funktion af s, at benytte sætningen om differentiation af omvendt afbildning).

Bemærk, at T–fordelingen med $f = 1$ er identisk med Cauchy–fordelingen (opgave 5.4.4).

(b) Vis at middelværdi og varians i T–fordelingen er givet ved

$$\mathrm{E}T = 0 \text{ for } f \geq 2,$$

$$\mathrm{var}(T) = \frac{f}{f-2} \text{ for } f \geq 3.$$

(Vink: Ifølge sætning 4.2.2 (c) er

$$\mathrm{E}(T^2) = f\mathrm{E}(U^2)\mathrm{E}\frac{1}{U_1^2+\cdots+U_f^2} = f\mathrm{E}(U^2)\mathrm{E}\frac{1}{2X}$$

hvor $2X$ er χ^2–fordelt med f frihedsgrader, dvs. X er Γ–fordelt med $\lambda = \frac{f}{2}$. Her er det let at udtrykke $\mathrm{E}\frac{1}{X}$ som forholdet mellem to værdier af Γ–funktionen. Benyt herefter Γ–funktionens funktionalligning).

Opgave 7.1.4*.

(a) Vis at F–fordelingen med (f_1, f_2) frihedsgrader har tæthed

$$p(v) = \frac{f_1^{\frac{f_1}{2}} f_2^{\frac{f_2}{2}} v^{\frac{f_1}{2}-1}}{\mathrm{B}\left(\frac{f_1}{2}, \frac{f_2}{2}\right)(f_2 + f_1 v)^{\frac{f_1+f_2}{2}}}.$$

(Vink: F–fordelingen fremkommer pr. definition ved bijektiv transformation af en B–fordeling).

(b) Vis at middelværdien i F–fordelingen er givet ved

$$\mathrm{E}V = \frac{f_2}{f_2 - 2}$$

for $f_2 > 2$. (Vink: En F–fordelt størrelse kan tænkes fremkommet på formen

$$V = \frac{X_1/f_1}{X_2/f_2}$$

hvor X_1 og X_2 er uafhængige, χ^2–fordelte med f_1 og f_2 frihedsgrader. Se vinket til (b) i foregående opgave).

Opgave 7.1.5. Lad $R_1, \ldots, R_n$ være stokastisk uafhængige, ligefordelte på enhedsintervallet. For et givet tal $p \in]0, 1[$, sæt $k_n = [pn]$. Vi kan da opfatte $R_{(k_n)}$ (= den k_n'te blandt de ordnede observationer, jvf. eksempel 6.2.6) som p–fraktilen i den empiriske fordeling bestemt ved observationerne $R_1, \ldots, R_n$ (jvf. opgave 5.4.5). Vis at der for ethvert $\epsilon > 0$ gælder

$$P\left(|R_{(k_n)} - p| > \epsilon\right) \to 0$$

for $n \to \infty$, dvs. "den empiriske p–fraktil konvergerer i sandsynlighed mod den teoretiske p–fraktil". (Vink: Variansen i fordelingen af $R_{(k_n)}$ fås af opgave 7.1.2, idet fordelingen ifølge eksempel 6.2.6 er en B–fordeling. Benyt Chebychev's ulighed).

7.2. Den flerdimensionale normale fordeling.

Ved den *normerede n–dimensionale normale fordeling* forstås simpelthen fordelingen af et sæt $U = (U_1, \ldots, U_n)$, bestående af n uafhængige, normeret normalfordelte variable på $\mathbf{R}$. En normal fordeling på $\mathbf{R}^n$ defineres herefter som fordelingen af en variabel af formen

$$X = \mu + CU \qquad (\mu \in \mathbf{R}^n, C \text{ en } n \times n\text{–matrix}).$$

Bemærk at definitionen er helt analog til definitionen af en normal fordeling på $\mathbf{R}$, idet positionsparameteren μ blot er blevet til en vektor, og skalaparameteren er blevet erstattet med en matrix (eller lineær afbildning) C.

Vi begynder med nogle resultater, som vedrører den normerede fordeling. I de statistiske anvendelser af χ^2–fordeling, F–fordeling og T–fordeling dukker disse fordelinger ikke op på helt så simpel måde, som sætning 7.1.4 antyder. Det er ganske vist rigtigt, at de altid kan tænkes fremkommet ud fra normeret normalfordelte variable $U_1, \ldots, U_n$ på den angivne måde, men disse vil igen være givet som linearkombinationer af de egentlige "observationer". I praksis kræver udledningen af disse fordelinger derfor yderligere et resultat. Vi giver en version af dette resultat, og illustrerer dets betydning ved at vise en sætning (7.2.2), som har direkte relation til en simpel statistisk model.

Først minder vi om nogle definitioner og resultater i forbindelse med lineær algebra og n–dimensional geometri. Idet vi som sædvanligt opfatter elementer i $\mathbf{R}^n$ som $n \times 1$ søjler, defineres det *indre produkt* af x og $y \in \mathbf{R}^n$ som tallet

$$(x|y) = x_1 y_1 + \cdots + x_n y_n = x'y.$$

Normen af x er givet ved

$$\|x\| = \sqrt{(x|x)} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

For $x, y \in \mathbf{R}^n$ er $\|x - y\|$ således den sædvanlige (Euklidiske) afstand mellem de to punkter x og y.

Et sæt $e_1, \ldots, e_n$ af vektorer er en *ortonormal basis* i $\mathbf{R}^n$, hvis der gælder

$$(e_i|e_j) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{ellers.} \end{cases}$$

En $n \times n$–matrix kaldes ortonormal, hvis dens søjler udgør en ortonormal basis. Der gælder

$$\begin{aligned} & M \text{ ortonormal} \\ \Longleftrightarrow & M' \text{ ortonormal} \\ \Longleftrightarrow & M'M = MM' = I \end{aligned}$$

(hvor I betegner enhedsmatricen). De ortonormale matricer kan også karakteriseres ved den egenskab, at de tilsvarende lineære afbildninger bevarer indre produkter:

$$(Mx|My) = (Mx)'My = x'M'My = x'y = (x|y).$$

Specielt er de *normbevarende*:

$$\|Mx\| = \|x\|.$$

SÆTNING 7.2.1. *Lad $e_1, \ldots, e_n$ være en ortonormal basis i $\mathbf{R}^n$, og lad $X = (X_1, \ldots, X_n)$ være et sæt af uafhængige, normeret normalfordelte variable. Definér*

$$U_i = (e_i|X) \qquad (i = 1, \ldots, n).$$

Da er $U_1, \ldots, U_n$ igen uafhængige, normeret normalfordelte.

Eller, i en lidt mere matrix–orienteret formulering:

Lad $X \in \mathbf{R}^n$ være normeret normalfordelt, og lad M være en ortonormal matrix. Da er $U = MX$ igen normeret normalfordelt.

Bemærk: For $n = 2$ har vi allerede konstateret dette resultats gyldighed i beviset for den normale fordelings foldningsegenskab (sætning 6.3.2); i dette tilfælde består en ortonormal basis jo netop af to vektorer af formen (α, β) og $(-\beta, \alpha)$ med $\alpha^2 + \beta^2 = 1$.

BEVIS. Vi anvender transformationssætningen for det lineære tilfælde, jvf. eksempel 6.2.4. Transformationen

$$x = (x_1, \ldots, x_n) \longrightarrow u = ((e_1|x), \ldots, (e_n|x))$$

kan på matrixform skrives

$$x \longrightarrow u = Mx,$$

hvor M' er den ortonormale matrix hvis søjler netop er $e_1, \ldots, e_n$. Determinanten for en sådan matrix er ± 1, og da M er normbevarende er tæthedens værdi i punktet $x = M^{-1}u$

$$\begin{aligned}
\left(\frac{1}{\sqrt{2\pi}}\right)^n & \exp\left(-\frac{1}{2}(x_1^2 + \cdots + x_n^2)\right) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\|x\|^2\right) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\|u\|^2\right) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}(u_1^2 + \cdots + u_n^2)\right).
\end{aligned}$$

Heraf følger sætningen umiddelbart.

SÆTNING 7.2.2. *Lad $Y_1, \ldots, Y_n$ være stokastisk uafhængige, identisk normalfordelte med middelværdi μ og varians σ^2. Definer gennemsnittet $\bar{Y}$ og kvadratafvigelsessummen* SSD *ved*

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n),$$
$$\mathrm{SSD} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Da er $\bar{Y}$ *og* SSD *stokastisk uafhængige,* $\bar{Y}$ *normalfordelt* $(\mu, \sigma^2/n)$ *og* SSD χ^2*–fordelt med* $n-1$ *frihedsgrader og skalaparameter* σ^2*. For* $\mu = 0$ *er størrelsen*

$$T = \frac{\sqrt{n}\bar{Y}}{\sqrt{\frac{1}{n-1}\mathrm{SSD}}}$$

T-fordelt med $n-1$ *frihedsgrader.*

Bemærkninger. I statistisk sammenhæng er $\bar{Y}$ det naturlige estimat for μ og $s^2 = \frac{1}{n-1}\mathrm{SSD}$ er det naturlige estimat for σ^2. T-størrelsen benyttes til test for hypotesen $\mu = 0$.

Bevis. Lad X betegne den n–dimensionale variable, hvis koordinater er de standardiserede variable $X_i = (Y_i - \mu)/\sigma$, $i = 1, \ldots, n$, som er uafhængige, normeret normalfordelte. Lad e_1 betegne enhedsvektoren $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$. Det éndimensionale underrum udspændt af denne vektor består af alle vektorer af formen $(t, \ldots, t)$. Det $n-1$–dimensionale ortogonale komplement til dette underrum udgøres af vektorerne $(x_1, \ldots, x_n)$ med koordinatsum $x_1 + \cdots + x_n = 0$. Lad $e_2, \ldots, e_n$ være en ortonormal basis for dette underrum. Så er $e_1, e_2, \ldots, e_n$ en ortonormal basis for $\mathbf{R}^n$. Sæt $U_i = (e_i|X)$. Da er

$$\begin{aligned}
\bar{Y} &= \mu + \sigma\bar{X} \\
&= \mu + \sigma\frac{1}{n}(X_1 + \cdots + X_n) \\
&= \mu + \sigma\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)\right) \\
&= \mu + \frac{\sigma}{\sqrt{n}}U_1,
\end{aligned}$$

$$\begin{aligned}
\mathrm{SSD} &= (Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2 \\
&= \sigma^2\left((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right) \\
&= \sigma^2(X_1^2 + \cdots + X_n^2 - n\bar{X}^2) \\
&= \sigma^2\left((U_1^2 + \cdots + U_n^2) - U_1^2\right) \\
&= \sigma^2(U_2^2 + \cdots + U_n^2)
\end{aligned}$$

og for $\mu = 0$ er

$$\begin{aligned}
T &= \frac{\sqrt{n}\bar{Y}}{\sqrt{\frac{1}{n-1}\mathrm{SSD}}} \\
&= \frac{\sigma\sqrt{n}\bar{X}}{\sigma\sqrt{\frac{1}{n-1}(U_2^2 + \cdots + U_n^2)}} \\
&= \frac{U_1}{\sqrt{(U_2^2 + \cdots + U_n^2)/(n-1)}}.
\end{aligned}$$

Efter disse omskrivninger følger sætningen umiddelbart af resultaterne fra foregående paragraf.

Den generelle n–dimensionale normale fordeling. Betragt en variabel af formen

$$X = \mu + CU$$

hvor U er normeret normalfordelt på $\mathbf{R}^n$, $\mu \in \mathbf{R}^n$, C en $n \times n$–matrix. Hvis den affine transformation $u \to \mu + Cu$ er bijektiv (dvs. C er regulær) kan vi opskrive tætheden for fordelingen af X (jvf. eksempel 6.2.4):

$$\begin{aligned} q(x) &= \frac{p\left(C^{-1}(x-\mu)\right)}{|\det(C)|} \\ &= \frac{1}{\sqrt{2\pi}^n |\det(C)|} \exp\left(-\frac{1}{2}\left(C^{-1}(x-\mu)\right)'\left(C^{-1}(x-\mu)\right)\right) \\ &= \frac{1}{\sqrt{2\pi}^n |\det(C)|} \exp\left(-\frac{1}{2}(x-\mu)'(C^{-1})'C^{-1}(x-\mu)\right) \\ &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \end{aligned}$$

hvor vi i sidste omskrivning har indført matricen $\Sigma = CC'$. Der gælder jo så

$$\left(C^{-1}\right)' C^{-1} = (C')^{-1}C^{-1} = (CC')^{-1} = \Sigma^{-1}.$$

Samtidig har vi omskrevet normeringsfaktoren ved hjælp af følgende udregning:

$$|\det(C)| = \sqrt{\det(C)^2} = \sqrt{\det(C)\det(C')} = \sqrt{\det(CC')} = \sqrt{\det(\Sigma)}.$$

Hermed har vi udtrykt den generelle flerdimensionale normalfordelings tæthed ved parametre μ og Σ, som svarer til middelværdi og varians i det éndimensionale tilfælde. Parameteren μ er fordelingens middelværdi i den forstand, at der koordinatvis gælder $\mathrm{E}X_i = \mu_i$. Matricen Σ er fordelingens *kovariansmatrix,* dvs. den $n \times n$–matrix hvis (i,j)'te element er kovariansen mellem X_i og X_j (og hvis diagonalelementer således er $\mathrm{var}(X_1), \ldots, \mathrm{var}(X_n)$). Denne sidste påstand er en elementær konsekvens af definitionen af et matrixprodukt. Vi vil senere bevise den for $n = 2$. Opskrivningen af beviset i det generelle tilfælde vil vi overlade til læseren, og vi vil i det hele taget ikke sige mere om den generelle, n–dimensionale normale fordeling, men i stedet gå lidt mere i detaljer med

Den todimensionale normale fordeling. Lad U_1 og U_2 være uafhængige, normeret normalfordelte. En todimensional normal variabel (X,Y) fås da ved

$$\begin{aligned} X &= \mu_X + c_{11}U_1 + c_{12}U_2, \\ Y &= \mu_Y + c_{21}U_1 + c_{22}U_2. \end{aligned}$$

Under antagelse af, at koefficientmatricen

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

er regulær kan vi opskrive fordelingens tæthed på samme måde som i det generelle tilfælde. Men først vil vi modificere konstruktionen lidt. Ifølge sætning 7.2.1 (og beviset for sætning 6.3.2) kan vi, uden at det får indflydelse på fordelingen af (X, Y), tænke os at U_1 og U_2 er fremkommet ved transformation af to andre uafhængige, normeret normalfordelte variable V_1 og V_2 på formen

$$\begin{aligned} U_1 &= \alpha V_1 + \beta V_2, \\ U_2 &= -\beta V_1 + \alpha V_2, \end{aligned}$$

hvor α og β er vilkårlige tal med kvadratsum 1. Ved at indsætte disse udtryk for U_1 og U_2 i konstruktionen af X og Y får vi fremstillingen

$$\begin{aligned} X &= \mu_X + (\alpha c_{11} - \beta c_{12})V_1 + (\beta c_{11} + \alpha c_{12})V_2, \\ Y &= \mu_Y + (\alpha c_{21} - \beta c_{22})V_1 + (\beta c_{21} + \alpha c_{22})V_2. \end{aligned}$$

Hermed har vi indset, at den samme todimensionale normalfordeling kan fremkomme på adskillige måder ved lineært affin transformation af en normeret fordeling. Dette svarer til et fænomen, man også kender i dimension 1, hvor skift af en skalaparameters fortegn (for en normal fordeling, eller mere generelt for en symmetrisk fordeling) ikke vil ændre fordelingen. Vi kan udnytte dette til at vælge koeefficientmatricen på en måde, som letter de følgende overvejelser. Ved at vælge (α, β) som en af de to enhedsvektorer, der er vinkelret på vektoren (c_{12}, c_{11}), opnår vi jo en fremstilling, hvor anden koefficient i første linie er 0. Med passende nye betegnelser kommer den herefter til at se sådan ud:

$$\begin{aligned} X &= \mu_X + aV_1, \\ Y &= \mu_Y + bV_1 + cV_2. \end{aligned}$$

Herefter kan vi glemme alt om den mere generelle fremstilling, vi begyndte med, idet enhver todimensional normalfordeling kan konstrueres på denne måde.

Vi indfører følgende betegnelser:

$$\begin{aligned} \sigma_X^2 &= \text{var}(X) = a^2, \\ \sigma_Y^2 &= \text{var}(Y) = b^2 + c^2, \\ \rho = \text{corr}(X, Y) &= \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{ab}{\sigma_X \sigma_Y} = \frac{ab}{\sqrt{a^2(b^2 + c^2)}}. \end{aligned}$$

Til udledning af fordelingens tæthed kan vi nu udnytte den udregning, vi allerede har gennemført i det generelle tilfælde. Vi indførte dér matricen $\Sigma = CC'$, som i dette tilfælde bliver

$$\Sigma = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Bemærk, at dette er fordelingens kovariansmatrix, i overensstemmelse med hvad der blev sagt om Σ i det generelle tilfælde. Inversion af denne 2×2–matrix efter en velkendt metode giver

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{\sigma_X^2\sigma_Y^2 - (\rho\sigma_X\sigma_Y)^2} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix} \\ &= \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_X^2} & \frac{-\rho}{\sigma_X\sigma_Y} \\ \frac{-\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{bmatrix}. \end{aligned}$$

Her må vi forudsætte $\rho^2 \neq 1$. Tilfældet $\rho = \pm 1$, hvor kovariansmatricen åbenbart er singulær, svarer til en fordeling som er koncentreret på en linie, jvf. sætning 4.4.1 (b).

Tætheden for (X, Y) fås herefter (ved at indsætte i formlen fra det n–dimensionale tilfælde) til

$$q(x,y) = \frac{\exp\left(-\frac{1}{2}\begin{bmatrix} x-\mu_X & y-\mu_Y \end{bmatrix}\Sigma^{-1}\begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}\right)}{\sqrt{(2\pi)^2\det(\Sigma)}}$$

$$= \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right)}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2(1-\rho^2)}}.$$

Niveaukurverne for denne funktion af to variable ses at være koncentriske, ligedannede ellipser med centrum i (μ_X, μ_Y). Disse ellipsers form og orientering i forhold til akserne bestemmes af parametrene σ_X^2, σ_Y^2 og ρ. For $\rho = 0$ fås åbenbart ellipser med hovedakser, som er parallelle med koordinatakserne. Forholdet mellem længden af den vandrette og den lodrette akse er σ_X/σ_Y. For $\rho \neq 0$ fås ellipser, som ligger skævt i forhold til koordinatsystemet. Hældningen af den længste hovedakse får samme fortegn som ρ, og hovedreglen er, at værdier af ρ nær ± 1 svarer til langstrakte ellipser, som ligger tydeligt skævt i forhold til akserne. I tilfældet $\sigma_X = \sigma_Y$ har den længste hovedakse hældning ± 1 (med samme fortegn som ρ). Ellipsens "fladtrykthed" er i dette tilfælde givet ved at forholdet mellem længste og korteste hovedakse er $\sqrt{\frac{1+|\rho|}{1-|\rho|}}$.

Korrelationens fortolkning som et mål for graden af samvariation mellem de to variable fremgår tydeligt af ovenstående. En alternativ beskrivelse af en sådan samvariation får man ved at se på, hvorledes den betingede fordeling af Y, givet $X = x$, afhænger af x.

Her bliver vi nødt til at indskyde en bemærkning om betingning i det kontinuerte tilfælde. Vi har tidligere konstateret, at definitionen af en betinget fordeling, givet en hændelse med positiv sandsynlighed, kan overføres stort set uændret fra det diskrete tilfælde (opgave 5.4.2 og 5.4.3). Det gælder også for flerdimensionale kontinuerte fordelinger. Men da hændelsen $\{X = x\}$ har sandsynlighed 0, må der en anden definition til i dette tilfælde. Man kan give gode argumenter for følgende definition: Hvis (X, Y) har en fordeling i planen med tæthed $p(x, y)$, så er den betingede fordeling af Y, givet $X = x$, givet ved tætheden

$$p(y|x) = \frac{p(x,y)}{\int p(x,y)\,dy}.$$

Altså: Tætheden for en sådan betinget fordeling får man ved at opfatte $p(x, y)$ som funktion af y (x fast) og normere den til en tæthed. Bemærk at integralet i nævneren netop er tætheden for den marginale fordeling af X, udregnet i punktet x, så vi får identiteten

$$p(x,y) = q(x)p(y|x)$$

hvor q er tætheden for fordelingen af X. Alene denne formel, som helt svarer til en simpel relation mellem sandsynlighedsfunktioner i det diskrete tilfælde (f.eks. for fordelinger på $\mathbf{Z} \times \mathbf{Z}$) er et godt argument for definitionen. Et andet argument er, at den foreslåede betingede fordeling fås som grænse for den (veldefinerede) betingede fordeling af Y, givet $X \in [x, x+h]$, for $h \to 0$.

Vi vender tilbage til den todimensionale normale fordeling. Fremstillingen af $p(x, y)$ på formen $q(x)p(y|x)$ fås her ved følgende omskrivning:

$$\frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right)}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2(1-\rho^2)}}$$

$$= \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} - \rho\frac{x-\mu_X}{\sigma_X}\right)^2 - \rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)\right)}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2(1-\rho^2)}}$$

$$= \frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(-\frac{1}{2\sigma_X^2}(x-\mu_X)^2\right) \times$$

$$\frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp\left(-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)\right)^2\right)$$

Eftersom første faktor i sidste udtryk netop er den marginale tæthed for fordelingen af X (normalfordelingen med middelværdi μ_X og varians σ_X^2), må den anden faktor som funktion af y være tætheden for den betingede fordeling af Y, givet $X = x$. Vi konkluderer, at denne betingede fordeling ligeledes er normal, med middelværdi $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ og varians $\sigma_Y^2(1 - \rho^2)$. Bemærk at denne varians ikke afhænger af x. For $\rho \neq 0$ er den mindre end variansen σ_Y^2 i den marginale fordeling af Y, hvilket intuitivt har den forklaring, at kendskab til X sætter os i stand til at forudsige Y mere præcist end det er muligt uden kendskab til X. Middelværdien $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ i den betingede fordeling kan fortolkes som det bedste gæt på værdien af Y, baseret på den observerede værdi x af X. Bemærk, at denne *betingede middelværdi* (som den kaldes) afhænger lineært af x. Dette er, ligesom den ovenfor bemærkede egenskab at variansen i den betingede fordeling er konstant, noget helt specielt for den todimensionale normale fordeling. Hældningen $\rho\frac{\sigma_Y}{\sigma_X}$ af den linie, der beskriver Y's betingede middelværdi som funktion af x, kaldes *regressionskoefficienten*. Bemærk at den har samme fortegn som ρ, og for $\sigma_X = \sigma_Y$ er den simpelthen lig med ρ. Heri ligger en klar bekræftelse af korrelationskoefficientens relevans som et mål for graden af samvariationen, i hvert fald for den todimensionale normale fordelings vedkommende.

Efter alle disse besværlige udregninger kan vi nu glæde os over at have nået en konklusion, som i virkeligheden fremgår umiddelbart af den oprindelige konstruktion af X og Y ud fra normerede variable V_1 og V_2. Hvis man, ved at regne baglæns, udtrykker de oprindelige koefficienter a, b og c ved parametrene σ_X, σ_Y og ρ får man nemlig (som én af løsningerne – de tre andre fremkommer ved passende skift af fortegn på a, b og c)

$$\begin{aligned} a &= \sigma_X, \\ b &= \rho\sigma_Y, \\ c &= \sqrt{1-\rho^2}\,\sigma_Y. \end{aligned}$$

Vi kan derfor skrive den oprindelige konstruktion på formen

$$\begin{aligned} X &= \mu_X + \sigma_X V_1, \\ Y &= \mu_Y + \rho\sigma_Y V_1 + \sqrt{1-\rho^2}\,\sigma_Y V_2 \\ &= \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) + \sqrt{1-\rho^2}\,\sigma_Y V_2. \end{aligned}$$

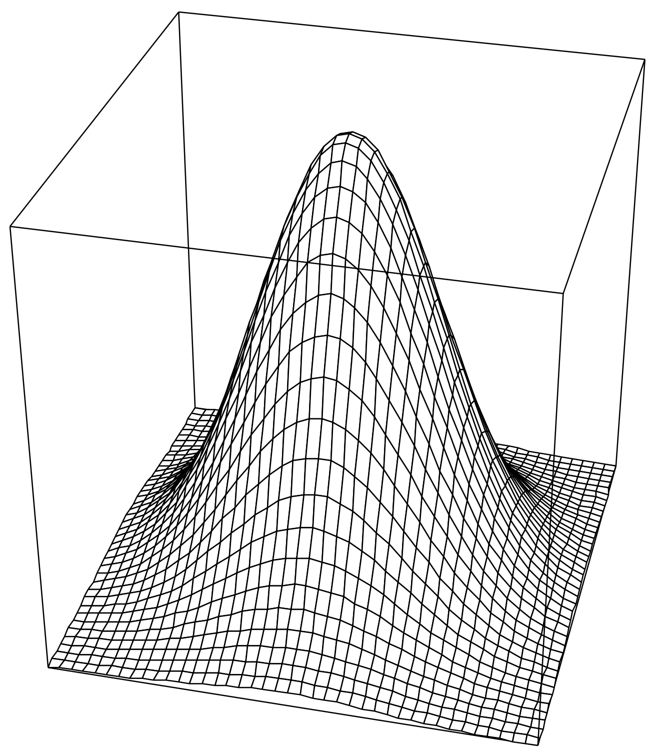
Af denne opskrift fremgår det tydeligt, at den eneste fornuftige forudsigelse af Y efter observation af X går ud på at opfatte Y som normalfordelt med middelværdi $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$ og varians $\sigma_Y^2(1-\rho^2)$. At kende X – eller V_1 – er jo det samme som at kende andet led i det sidste udtryk for Y. Og sidste led kan vi ikke forudsige på anden måde end ved at angive dets fordeling, eftersom V_2 er stokastisk uafhængig af det vi har observeret.

OPGAVE 7.2.1. Vis at *regressionslinien* (som den kaldes) i den todimensionale normale fordeling

$$y = \mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x - \mu_X)$$

kan karakteriseres som den linie gennem (μ_X, μ_Y) der skærer konturellipserne i de punkter hvor de har lodret tangent.

OPGAVE 7.2.2. Vis for den todimensionale normale fordeling med $\sigma_X = \sigma_Y$ at forholdet mellem længste og korteste hovedakse for konturellipserne er $\sqrt{\frac{1+|\rho|}{1-|\rho|}}$. (Vink: En sådan fordeling kan konstrueres ved en 45^o rotation af fordelingen af (aU_1, bU_2)).



Tætheden for den normerede todimensionale normalfordeling