



Föreläsning 02

Statistisk inläring och dataanalys (Kungliga Tekniska Högskolan)

Föreläsning 2

2.1 Flerdimensionella fördelningar

I många verkliga dataanalytiska situationer är vi intresserade av flera beroende stokastiska variabler X_1, \dots, X_m , istället för bara en variabel X . Det kan därmed vara användbart att ha modeller som kan uttrycka sannolikheten av att se *simultan utfall* liksom $\mathbf{x} = (x_1, \dots, x_m)^T$. Ett sådant utfall är en observation av den m -dimensionella *stokastiska vektorn* $\mathbf{X} = (X_1, \dots, X_m)^T$.

Exempel 2.1. Låt X_1 beteckna den stokastiska variabeln med utfall $\{0, 1\}$, där 1 betecknar att det har varit inbrott i ett visst hus och 0 betecknar att inget inbrott har inträffat där. Låt X_2 beteckna stokastiska variabeln med utfall $\{0, 1\}$, där 1 betecknar att polisen kommer till huset och 0 betecknar att de inte kommer. Vi skulle ha nytta av en förståelse av de simultana sannolikheterna:

$$P(0, 0), P(0, 1), P(1, 0), \text{ och } P(1, 1),$$

där $P(x_1, x_2)$ betecknar sannolikheten av det simultana utfallet (x_1, x_2) av $\mathbf{X} = (X_1, X_2)^T$.

I Exempel 1.12, utfallsrummet är ändligt och därför är $\mathbf{X} = (X_1, X_2)^T$ ett exempel av en *diskret stokastisk vektor*; dvs, en stokastisk vektor som tar värden i en uppräknelig mängd. En sådan stokastisk vektor har *simultan sannolikhetsfunktion*

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, \dots, x_m) = P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m).$$

Om \mathbf{X} tar istället värden i en ouppräknelig mängd kallas det en *kontinuerlig stokastisk vektor*. En kontinuerlig stokastisk vektor $\mathbf{X} = (X_1, \dots, X_m)^T$ har *simultan täthetsfunktion*

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, \dots, x_m),$$

där

$$\int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_m} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = P(X_1 \leq a_1, \dots, X_m \leq a_m),$$

för alla $a_1, \dots, a_m \in \mathbb{R}$.

Under denna kurs betecknar vi vektorer med fetstilt eller understruken bokstäver: $\mathbf{X} = (X_1, \dots, X_m)^T$. För ett stickprov $\mathbf{X}_1, \dots, \mathbf{X}_n$ från fördelningen av \mathbf{X} skriver vi $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,m})^T$ för alla $i = 1, \dots, n$. Så $X_{i,j}$ är utfallet av X_j för det i :te elementet \mathbf{X}_i i stickprovet.

Om $S = \{i_1, i_2, \dots, i_k\}$ är en delmängd av $\{1, 2, \dots, m\}$ kan vi betrakta också den stokastiska *delvektorn* $\mathbf{X}_S = (X_{i_1}, \dots, X_{i_k})^T$ av $\mathbf{X} = (X_1, \dots, X_m)^T$. Dessa har utfall $\mathbf{x}_S = (x_{i_1}, \dots, x_{i_k})^T \in \mathbb{R}^k$. Vi är nu intresserade av fördelningen för \mathbf{X}_S , dvs. hur sannolika olika värden på \mathbf{X}_S är oavsett vad som händer med de resterande komponenterna i \mathbf{X} .

Denna fördelning kallas för *marginalfördelningen* av \mathbf{X}_S . I det diskreta fallet har vi den *marginella sannolikhetsfunktionen*

$$f_{\mathbf{X}_S}(\mathbf{x}_S) = \sum_{\mathbf{x}_{\{1,2,\dots,m\} \setminus S}} f_{\mathbf{X}}(\mathbf{x}),$$

dvs. vi summerar den simultana täthetsfunktionen $f_{\mathbf{X}}(\mathbf{x})$ över alla möjliga värden på \mathbf{x} där \mathbf{x}_S hålls fixt. (För enkelhets skull kommer vi inte att skriva ner gränserna i summor för sannolikhetsfunktioner när dessa kan härledas från sannolikhetsfunktionens stöd.) För kontinuerliga stokastiska vektorer har vi på samma sätt den *marginella täthetsfunktionen*

$$f_{\mathbf{X}_S}(\mathbf{x}_S) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_{\{1,2,\dots,m\} \setminus S}.$$

När vi beräknar dessa fördelningar säger vi att vi *marginaliserar* ut de variabler som inte är i S .

Det enklaste fallet är då $m = 2$. Om vi tar $S = \{1\}$ får vi

$$f_{X_1}(x_1) = \sum_{x_2} f_{X_1, X_2}(x_1, x_2),$$

i det diskreta fallet och

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1, X_2}(x_1, x_2) dx_2$$

i det kontinuerliga fallet. Dessa formler är desamma som vi såg i föregående kurs och motsvarar marginalfördelningarna i det tvådimensionella fallet.

Exempel 2.2. Vi fortsätter med Exempel 1.12. Om vi skulle vilja veta den marginella sannolikheten av $X_2 = 0$ räknar vi

$$P(X_2 = 0) = f_{X_2}(0) = f_{\mathbf{X}}(0, 0) + f_{\mathbf{X}}(1, 0) = P(0, 0) + P(1, 0);$$

dvs, vi summerar sannolikheterna av $(0, 0)$ och $(1, 0)$. Resultatet är sannolikheten att polisen kommer till huset oavsett om något inbrott har skett.

Vi avslutar denna föreläsning med några viktiga exempel på flerdimensionella fördelningar.

2.1.1 Multinomialfördelningen

Låt $\mathbf{X} = (X_1, \dots, X_m)^T$ vara en stokastisk vektor där X_i har möjliga utfall $\{0, 1, \dots, n\}$ för alla $i = 1, \dots, m$. För ett positivt heltal n och $0 \leq p_1, \dots, p_m \leq 1$ antar vi att $x_1 + \dots + x_m = n$ och $p_1 + \dots + p_m = 1$. Vi säger \mathbf{X} är *multinomialfördelad* om

$$f_{\mathbf{X}}(x_1, \dots, x_m) = \frac{n!}{x_1! \dots x_m!} p_1^{x_1} \dots p_m^{x_m}.$$

Vi skriver detta som $\mathbf{X} \sim \text{Multi}(n, p_1, \dots, p_m)$.

Multinomialfördelningen modellerar en situation där ett experiment med möjliga utfall $1, 2, \dots, m$ vilka inträffar med sannolikheter p_1, p_2, \dots, p_m . Detta experiment genomförs n gånger och X_i betecknar antalet gånger utfallet i inträffar. Man kan visa att marginalfördelningen $f_{X_i}(x_i)$ för alla $i = 1, 2, \dots, m$ är en binomialfördelning (se övning 2); dvs. $X_i \sim \text{Bin}(n, p_i)$ om $\mathbf{X} \sim \text{Multi}(n, p_1, \dots, p_m)$.

Eftersom summan av x_1, x_2, \dots, x_m alltid är lika med n räcker det att sätta x_1, x_2, \dots, x_{m-1} och ta $x_m = n - x_1 - \dots - x_{m-1}$ för att specificera ett visst utfall av \mathbf{X} . På samma sätt har vi att $p_m = 1 - p_1 - \dots - p_{m-1}$. En alternativ definition av multinomialfördelningen är då som en $(m-1)$ -dimensionell fördelning med sannolikhetsfunktion

$$\begin{aligned} & f_{X_1, \dots, X_{m-1}}(x_1, \dots, x_{m-1}) \\ &= \frac{n!}{x_1! \dots x_{m-1}! \left(m - \sum_{i=1}^{m-1} x_i\right)!} p_1^{x_1} \dots p_{m-1}^{x_{m-1}} \left(1 - \sum_{i=1}^{m-1} p_i\right)^{m - \sum_{i=1}^{m-1} x_i}. \end{aligned}$$

Denna mindre elegantare variant har fördelen att fallet $m = 2$ ger (den endimensionella) binomialfördelningen. I denna kurs kommer vi hålla oss till den första definitionen om inget annat anges.

2.1.2 Flerdimensionella normalfördelningen

Fixa parametrarna $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma = [\Sigma_{i,j}]_{i,j=1}^m \in \mathbb{R}^d \times \mathbb{R}^m$ en positivt definit matris (dvs, en symmetrisk matris med bara positiva egenvärden). Vi säger $\mathbf{X} = (X_1, \dots, X_m)^T$ har en *flerdimensionell normalfördelning* (eller helt enkelt att \mathbf{X} är normalfördelad) med parametrar $\boldsymbol{\mu}$ och Σ om \mathbf{X} har täthetsfunktion

$$f_{\mathbf{X}}(\mathbf{x} | \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{m/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{x})^T \Sigma^{-1}(\boldsymbol{\mu} - \mathbf{x})\right),$$

där $\det(\Sigma)$ är determinanten av Σ (positiv eftersom Σ är positivt definit). Den flerdimensionella normalfördelningen generaliserar den endimensionella normalfördelningen och de sammanfaller för $m = 1$. Vi skriver $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ om \mathbf{X} är normalfördelad med parametrar $\boldsymbol{\mu}$ och Σ . (Man kan också definiera normalfördelningar för Σ som inte är positivt definita utan positivt *semidefinita* – med egenvärden som tillåts vara noll – men i dessa har ingen väldefinierad täthetsfunktion då Σ inte längre är inverterbar.)

Precis som dess endimensionella kusin har den flerdimensionella normalfördelningen ett antal användbara egenskaper.

Sats 2.1. Låt $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ för någon $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma \in \mathbb{R}^{m \times m}$ (positivt definit). Om $S \subset \{1, \dots, m\}$ har vi att delvektorn \mathbf{X}_S uppfyller

$$\mathbf{X}_S \sim N(\boldsymbol{\mu}_S, \Sigma_{SS})$$

där $\Sigma_{SS} = [\Sigma_{i,j}]_{i,j \in S}$ (delmatrisen av Σ med rader och kolumner givna av S).

Bevis. Antag utan inskränkning att $S = \{1, \dots, k\}$ så att vi har $\mathbf{X}_S = [X_1, X_2, \dots, X_k]^T$ samt $\mathbf{x}_S = [x_1, x_2, \dots, x_k]^T$. Vidare sätter vi $T = \{1, 2, \dots, d\} \setminus S = \{k+1, k+2, \dots, m\}$ så att vi kan skriva $\mathbf{X} = [\mathbf{X}_S^T \mathbf{X}_T^T]^T$ samt $\mathbf{x} = [\mathbf{x}_S^T \mathbf{x}_T^T]^T$.

För att beräkna den marginella täthetsfunktionen $f_{\mathbf{X}_S}(\mathbf{x}_S)$ för \mathbf{X}_S måste vi ta den simultana täthetsfunktionen $f_{\mathbf{X}}(\mathbf{x})$ och integrera denna med avseende på \mathbf{x}_T (dvs. marginalisera bort \mathbf{X}_T). Svårigheten med detta är att täthetsfunktionen $f_{\mathbf{X}}(\mathbf{x})$ involverar \mathbf{x}_S och \mathbf{x}_T på ett sätt som gör det svårt att integrera med avseende på den ena men inte den andra. Mer specifikt så har vi följande uttryck i exponenten

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}).$$

Om vi delar upp $\boldsymbol{\mu}$ på samma sätt som \mathbf{x} , dvs. vi tar $\boldsymbol{\mu} = [\boldsymbol{\mu}_S^T \boldsymbol{\mu}_T^T]^T$ så kan vi skriva

$$\begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T \end{bmatrix}$$

men matrismultiplikationen med Σ^{-1} ställer till det eftersom den blandar termer som beror på \mathbf{x}_S med termer som beror på \mathbf{x}_T . Vi skulle behöva ett sätt att skriva om Σ^{-1} som separerar dessa termer.

För Σ själv är inte detta ett större problem. Vi kan göra en blockuppdelning med avseende på S och T vilket ger

$$\Sigma = \begin{bmatrix} \Sigma_{SS} & \Sigma_{ST} \\ \Sigma_{TS} & \Sigma_{TT} \end{bmatrix}$$

där $\Sigma_{SS} = [\Sigma_{i,j}]_{i,j \in S}$, $\Sigma_{ST} = [\Sigma_{i,j}]_{i \in S, j \in T}$, $\Sigma_{TS} = \Sigma_{ST}^T$ och $\Sigma_{TT} = [\Sigma_{i,j}]_{i,j \in T}$. Om vi multiplicerar Σ med \mathbf{x} kan vi med denna blockuppdelning separera termerna som beror på \mathbf{x}_S från de som beror på \mathbf{x}_T . Frågan är nu hur vi kan hitta en liknande blockuppdelning för Σ^{-1} .

För att invertera en 2×2 -blockmatris på formen

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

använder man Schurkomplementet av A i M , vilket ges av

$$M/A = D - CA^{-1}B.$$

Inversen av M^{-1} ges då av

$$M^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (M/A)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}.$$

I vårt fall har vi $A = \Sigma_{SS}$, $B = \Sigma_{ST}$, $C = \Sigma_{TS}$ och $D = \Sigma_{TT}$, vilket ger

$$\Sigma^{-1} = \begin{bmatrix} I & -\Sigma_{SS}^{-1}\Sigma_{ST} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{SS}^{-1} & 0 \\ 0 & (\Sigma/\Sigma_{SS})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Sigma_{TS}\Sigma_{SS}^{-1} & I \end{bmatrix}$$

där $\Sigma/\Sigma_{SS} = \Sigma_{TT} - \Sigma_{TS}\Sigma_{SS}^{-1}\Sigma_{ST}$. Eftersom Σ_{SS} är symmetrisk och $\Sigma_{ST}^T = \Sigma_{TS}$ kan vi skriva om detta som

$$\Sigma^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{TS}\Sigma_{SS}^{-1} & I \end{bmatrix}^T \begin{bmatrix} \Sigma_{SS}^{-1} & 0 \\ 0 & (\Sigma/\Sigma_{SS})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Sigma_{TS}\Sigma_{SS}^{-1} & I \end{bmatrix}.$$

Vi får då

$$\begin{aligned}
& (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\
&= \begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T \end{bmatrix}^T \begin{bmatrix} \mathbf{I} & 0 \\ -\Sigma_{TS} \Sigma_{SS}^{-1} & \mathbf{I} \end{bmatrix}^T \begin{bmatrix} \Sigma_{SS}^{-1} & 0 \\ 0 & (\Sigma / \Sigma_{SS})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\Sigma_{TS} \Sigma_{SS}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S) \end{bmatrix}^T \begin{bmatrix} \Sigma_{SS}^{-1} & 0 \\ 0 & (\Sigma / \Sigma_{SS})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_S - \boldsymbol{\mu}_S \\ \mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S) \end{bmatrix} \\
&= (\mathbf{x}_S - \boldsymbol{\mu}_S)^T \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S) \\
&\quad + (\mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S))^T (\Sigma / \Sigma_{SS})^{-1} (\mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S)).
\end{aligned}$$

Utöver detta har vi att

$$\det \Sigma = \det \Sigma_{SS} \cdot \det(\Sigma / \Sigma_{SS}).$$

Tillsammans ger detta

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{(2\pi)^{k/2} \sqrt{\det \Sigma_{SS}}} e^{-\frac{1}{2} (\mathbf{x}_S - \boldsymbol{\mu}_S)^T \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S)} \\
&\quad \frac{1}{(2\pi)^{(d-k)/2} \sqrt{\det(\Sigma / \Sigma_{SS})}} e^{-\frac{1}{2} (\mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S))^T (\Sigma / \Sigma_{SS})^{-1} (\mathbf{x}_T - \boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S))}.
\end{aligned}$$

Om vi nu integrerar med avseende på \mathbf{x}_T så försvinner den andra faktorn eftersom den är täthetsfunktionen för $N(\boldsymbol{\mu}_T - \Sigma_{TS} \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S), \Sigma / \Sigma_{SS})$ och därför integrerar till ett. Vi får då

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} \sqrt{\det \Sigma_{SS}}} e^{-\frac{1}{2} (\mathbf{x}_S - \boldsymbol{\mu}_S)^T \Sigma_{SS}^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S)},$$

dvs. att $\mathbf{X} \sim N(\boldsymbol{\mu}_S, \Sigma_{SS})$, vilket skulle visas. □

En konsekvens av denna sats är att $X_i \sim N(\mu_i, \Sigma_{i,i})$ så $E[X_i] = \mu_i$ och $\text{Var}[X_i] = \Sigma_{i,i}$.

Exempel 2.3. Låt $\mathbf{X} = (X_1, X_2, X_3)^T \sim N(\boldsymbol{\mu}, \Sigma)$ där

$$\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 23 \\ -4.5 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 14 & 32 & 50 \\ 32 & 77 & 122 \\ 50 & 122 & 194 \end{bmatrix}.$$

Det följer att $X_1 \sim N(6, 14)$ och $\mathbf{X}_{\{1,3\}} = (X_1, X_3)^T \sim N(\boldsymbol{\mu}_{\{1,3\}}, \Sigma_{\{1,3\}, \{1,3\}})$, där

$$\boldsymbol{\mu}_{\{1,3\}} = \begin{bmatrix} 6 \\ -4.5 \end{bmatrix}, \quad \Sigma_{\{1,3\}, \{1,3\}} = \begin{bmatrix} 14 & 50 \\ 50 & 194 \end{bmatrix}.$$

2.1.3 Flerdimensionella exponentialfamiljer

Precis som i det endimensionella fallet kan familjer av flerdimensionella distributioner bilda en exponentialfamilj. För en m -dimensionell fördelning med parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ kräver vi då att $f(\mathbf{x} \mid \boldsymbol{\theta})$ kan skrivas som

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = h(\mathbf{x}) c(\boldsymbol{\theta}) \exp \left(\sum_{j=1}^k w_j(\boldsymbol{\theta})^T t_j(\mathbf{x}) \right)$$

där $h : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$, $c : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ och $w_i : \mathbb{R}^d \rightarrow \mathbb{R}^{k_i}$ samt $t_i : \mathbb{R}^m \rightarrow \mathbb{R}^{k_i}$ för något positivt heltal k_i för alla $i = 1, 2, \dots, k$. Skillnaden är alltså att vi har $\mathbf{x} \in \mathbb{R}^m$ istället för $x \in \mathbb{R}$ samt att vi tillåter w_i och t_i att ta värden i \mathbb{R}^{k_i} och bildar skalärprodukten $w_i(\boldsymbol{\theta})^T t_i(\mathbf{x})$ (det senare är bara till för att förenkla notation och spelar ingen större roll). Annars betar sig dessa exponentialfamiljer likadant som i det endimensionella fallet.

Exempel 2.4. Likt binomialfördelningen bildar multinomialfördelningen $\text{Multi}(n, p_1, p_2, \dots, p_m)$ också en exponentialfamilj för fixt antal försök n med parameter $\boldsymbol{\theta} = [p_1, p_2, \dots, p_m]^T$. Vi ser detta genom att skriva om sannolikhetsfunktionen

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = \frac{n!}{x_1!x_2!\cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m} = \frac{n!}{x_1!x_2!\cdots x_m!} \exp\left(\sum_{i=1}^m x_i \log p_i\right) = h(\mathbf{x})c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^m w_i(\boldsymbol{\theta})t_i(\mathbf{x})\right)$$

där $h(\mathbf{x}) = n!/x_1!x_2!\cdots x_m!$, $c(\boldsymbol{\theta}) = 1$ och $w_i(\boldsymbol{\theta}) = \log p_i$ samt $t_i(\mathbf{x}) = x_i$ för alla $i = 1, 2, \dots, m$. Familjen av multinomialfördelningar för fixt n är också en fullständig exponentialfamilj (då antalet naturliga parametrar är lika med dimensionen på parameterutrymmet).

Exempel 2.5. För den flerdimensionella normalfördelningen med parametrar $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma \in \mathbb{R}^{m \times m}$ (positivt definit) har vi täthetsfunktionen

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

Vi kan nu skriva

$$\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x} - \boldsymbol{\mu}^T \Sigma^{-1}\mathbf{x} + \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu}.$$

Eftersom $\mathbf{x}^T \Sigma^{-1}\mathbf{x}$ är en skalär är den lika med spåret

$$\text{tr}(\mathbf{x}^T \Sigma^{-1}\mathbf{x}) = \text{tr}(\Sigma^{-1}\mathbf{x}\mathbf{x}^T)$$

där vi har utnyttjat att spåret är invariant med avseende på cykliska permutationer (dvs. att $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$). Vidare har vi att $\text{tr}(A^T B) = \text{vec}(A)^T \text{vec}(B)$ där $\text{vec}(A)$ avser *vektoriseringen* av A , dvs. om $A = [a_{ij}]_{i,j=1}^{n,k}$ så sätter vi att $\text{vec}(A)$ vektorn i \mathbb{R}^{nm} vars element i index $(i-1)n + j$ är lika med a_{ij} för $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$. Med hjälp av dessa transformationer har vi

$$\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{2}\text{vec}(\Sigma^{-1T} \text{vec} \mathbf{x}\mathbf{x}^T - (\Sigma^{-1}\boldsymbol{\mu})^T \mathbf{x} + \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu}).$$

Alltså kan vi skriva

$$f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})c(\boldsymbol{\theta}) \exp((w_1(\boldsymbol{\theta})^T t_1(\mathbf{x}) + w_2(\boldsymbol{\theta})^T t_2(\mathbf{x})),$$

där

$$\begin{aligned} h(\mathbf{x}) &= \frac{1}{(2\pi)^{m/2}} \\ c(\boldsymbol{\theta}) &= \exp\left(-\frac{1}{2}\log \det \Sigma - \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu}\right) \\ w_1(\boldsymbol{\theta}) &= \Sigma^{-1}\boldsymbol{\mu} \\ t_1(\mathbf{x}) &= \mathbf{x} \\ w_2(\boldsymbol{\theta}) &= -\frac{1}{2}\text{vec}(\Sigma^{-1}) \\ t_2(\mathbf{x}) &= \text{vec}(\mathbf{x}\mathbf{x}^T) \end{aligned}$$

Familjen av flerdimensionella normalfördelningar bildar således en exponentialfamilj.

2.2 Transformation av stokastiska variabler

Ett vanligt problem inom sannolikhets teorin är hur fördelningen av olika stokastiska variabler förändras när det avbildas genom olika funktioner. Vi har kanske en stokastisk variabel X och bildar $U = g(X)$ där $g: \mathbb{R} \rightarrow \mathbb{R}$. Om vi vet fördelningen för X kan vi använda denna, tillsammans med funktionen g , för att bestämma fördelningen för U .

Det första vi måste tänka på är stödet för fördelningen. Om vi har

$$\mathcal{A} = \{x \in \mathbb{R} : f_X(x) > 0\}$$

så innehåller \mathcal{A} alla värden som X kan anta (med sannolikhet ett). På samma sätt får vi då mängden av möjliga värden på U som

$$\mathcal{B} = \{u \in \mathbb{R} : u = g(x) \text{ för något } x \in \mathcal{A}\}.$$

Vi söker då för en täthetsfunktion eller sannolikhetsfunktion $f_U(u)$ med stöd \mathcal{B} .

I det diskreta fallet är situationen relativt enkel. Som vi sett i den tidigare kursen handlar det om att bestämma sannolikheten att X befinner sig i

$$g^{-1}(u) = \{x : g(x) = u\},$$

den så kallade *urbilden* av u via g (eng. "preimage"). Vi har då

$$f_U(u) = P(U = u) = P(g(X) = u) = P(X \in g^{-1}(u)) = \sum_{x \in g^{-1}(u)} f_X(x)$$

för $u \in \mathcal{B}$. Situationen är densamma för flerdimensionella stokastiska variabler $\mathbf{X} \in \mathbb{R}^n$ och $\mathbf{U} \in \mathbb{R}^n$ där $\mathbf{U} = g(\mathbf{X})$ för någon $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Där har vi

$$\mathcal{A} = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x}) > 0\} \quad \mathcal{B} = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u} = g(\mathbf{x}) \text{ för något } \mathbf{x} \in \mathcal{A}\}$$

och sannolikhetsfunktionen

$$f_U(\mathbf{u}) = \sum_{\mathbf{x} \in g^{-1}(\mathbf{u})} f_{\mathbf{X}}(\mathbf{x})$$

för $u \in \mathcal{B}$.

I det kontinuerliga fallet blir situationen mer komplex. Vi såg i den tidigare kursen att för en funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ så har vi

$$f_U(u) = f_X(g^{-1}(u)) \left| \frac{dg^{-1}}{du}(u) \right|, \quad (2.1)$$

för u i \mathcal{B} givet att g är injektiv på \mathcal{A} . (Notera att här så syftar g^{-1} på inversen av g eftersom denna är inverterbar på \mathcal{B} .)

Detta kan generaliseras till flera dimensioner. Låt oss börja med det tvådimensionella fallet. Då har vi de stokastiska variablerna X och Y samt en funktion $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ (injektiv på \mathcal{A}) och söker fördelningen för $(U, V) = g(X, Y)$. Man kan då visa att den simultana täthetsfunktionen för (U, V) ges av

$$f_{U,V}(u, v) = f_{X,Y}(g^{-1}(u, v)) |J|$$

för (u, v) i \mathcal{B} där vi har använt absolutvärdet av jacobideterminanten

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u}.$$

Detta värde kvantifierar hur pass mycket vår transformation $g(x, y)$ trycker ihop eller expanderar sannolikstätheten i en given punkt. Notera att detta är determinanten av jacobimatrisen för $g^{-1}(u, v)$, och generaliserar alltså på ett naturligt sätt skalfaktorn i (2.1).

Exempel 2.6. Man kan använda denna formel för att bestämma olika täthetsfunktioner, som till exempel täthetsfunktionen för summan eller differensen mellan två stokastiska variabler. Låt $X \sim N(0, 1)$ och $Y \sim N(0, 1)$ vara två oberoende stokastiska variabler. I detta fall har vi $\mathcal{A} = \mathbb{R}^2$ och $\mathcal{B} = \mathbb{R}^2$. Vi tar då $U = X + Y, V = X - Y$ och får

$$\begin{cases} u = x + y \\ v = x - y \end{cases} \Rightarrow \begin{cases} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{cases}.$$

Jacobideterminanten fås då som

$$J = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{2}$$

och vi har då

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) |J| \\ &= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{u+v}{2}\right)^2/2} e^{-\left(\frac{u-v}{2}\right)^2/2} \frac{1}{2} \\ &= \left(\frac{1}{\sqrt{2\pi} \cdot 2} e^{-u^2/4}\right) \left(\frac{1}{\sqrt{2\pi} \cdot 2} e^{-v^2/4}\right). \end{aligned}$$

Med andra ord är U och V två oberoende $N(0, 2)$ -fördelade stokastiska variabler.

I föregående kurs lärde vi oss *faltningsformeln* för täthetsfunktionen hos summan av två stokastiska variabler. Vi kunde så klart också använt denna i exemplet ovan för att hitta täthetsfunktionen för $U = X + Y$, men detta skulle inte gett oss täthetsfunktionen för $V = X - Y$ (och inte heller att U och V är oberoende). Faltningsformeln kan faktiskt ses som ett specialfall av transformationsformeln ovan (ta $U = X + Y$ och $V = Y$ och marginalisera med avseende på V för att få marginalfördelningen för U), men är mycket mer generell.

För fler än två dimensioner har formlerna samma form. Vi antar att vi har en stokastisk vektor $\mathbf{X} = [X_1, X_2, \dots, X_m]^T$ och en funktion $g: \mathbb{R}^m \rightarrow \mathbb{R}^m$ (injektiv på \mathcal{A}). Om vi tar

$$\mathcal{A} = \{\mathbf{x} \in \mathbb{R}^m : f_{\mathbf{X}}(\mathbf{x}) > 0\} \quad \mathcal{B} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = g(\mathbf{x}) \text{ för något } \mathbf{x} \in \mathcal{A}\}$$

så är $\mathbf{U} = g(\mathbf{X})$ en stokastisk vektor med värden i \mathcal{B} och vars täthetsfunktion ges av

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{X}}(g^{-1}(\mathbf{u})) |J|$$

där vi har jacobideterminanten

$$J = \frac{\partial(x_1, x_2, \dots, x_m)}{\partial(u_1, u_2, \dots, u_m)} = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_m} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_2}{\partial u_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_m}{\partial u_1} & \frac{\partial x_m}{\partial u_2} & \dots & \frac{\partial x_m}{\partial u_m} \end{vmatrix}.$$

Exempel 2.7. Låt oss anta att X_1, X_2, X_3, X_4 är oberoende och $\text{Exp}(1)$ -fördelade stokastiska variabler. Vi låter nu Y_1 vara det minsta av dessa fyra värden, Y_2 vara det näst minsta, Y_3 vara det näst största och Y_4 vara det största värdet. Dessa Y_1, Y_2, Y_3, Y_4 bildar då en stokastisk vektor och man kan visa (med hjälp av *ordningsstatistikor*, se övning 2) att de har den simultana täthetsfunktionen

$$f_{\mathbf{Y}}(\mathbf{y}) = 24e^{-y_1 - y_2 - y_3 - y_4}$$

för $0 \leq y_1 \leq y_2 \leq y_3 \leq y_4$. Antag att vi tar $U_1 = Y_1$, $U_2 = Y_2 - Y_1$, $U_3 = Y_3 - Y_2$ och $U_4 = Y_4 - Y_3$, dvs. vi mäter avståndet mellan de intilliggande värdena på X_1, X_2, X_3, X_4 . Om vi inverterar transformationen får vi

$$\begin{aligned} Y_1 &= U_1 \\ Y_2 &= U_1 + U_2 \\ Y_3 &= U_1 + U_2 + U_3 \\ Y_4 &= U_1 + U_2 + U_3 + U_4, \end{aligned}$$

och jacobideterminanten fås till

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 1,$$

vilket ger

$$f_{\mathbf{U}}(\mathbf{u}) = 24e^{-4u_1 - 3u_2 - 2u_3 - u_4}.$$

Med andra ord har vi att U_1, U_2, U_3, U_4 är oberoende exponentialfördelade stokastiska variabler med parametrar $1/4, 1/3, 1/2$ och 1 .

Vi kan också använda detta verktyg för att bevisa en av de viktigaste egenskaperna hos den flerdimensionella normalfördelningen: alla linjärkombinationer av en normalfördelad stokastisk vektor också är normalfördelade.

Sats 2.2. Låt $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ för någon $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma \in \mathbb{R}^{m \times m}$ (positivt definit). Vidare låt $A \in \mathbb{R}^{k \times m}$ vara en matris av full rang så att $k \leq m$ och låt $\mathbf{b} \in \mathbb{R}^k$. Då har vi att

$$A\mathbf{X} + \mathbf{b} \sim N(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$$

Bevis. Vi antar först att $k = m$, dvs. att A är en kvadratisk matris. Om vi sätter $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ har vi då att

$$\mathbf{X} = A^{-1}(\mathbf{Y} - \mathbf{b}) \quad \text{och} \quad \mathbf{x} = A^{-1}(\mathbf{y} - \mathbf{b})$$

eftersom A är av full rang. Avbildningen $\mathbf{y} \mapsto A^{-1}(\mathbf{y} - \mathbf{b})$ är linjär, så dess jacobimatrix är A^{-1} , vilket ger

$$J = \frac{\partial(x_1, x_2, \dots, x_m)}{\partial(y_1, y_2, \dots, y_m)} = \det A^{-1} = \frac{1}{\det A}.$$

Givet täthetsfunktionen för \mathbf{X} ,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})},$$

får vi då

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(A^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu})^T \Sigma^{-1}(A^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu})} \cdot \frac{1}{\det A}.$$

Eftersom

$$\begin{aligned} & (A^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu})^T \Sigma^{-1}(A^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}) \\ &= (A^{-1}(\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b}))^T \Sigma^{-1}(A^{-1}(\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b})) \\ &= (\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b})^T A^{-T} \Sigma^{-1} A^{-1} (\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b}) \\ &= (\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b})^T (A\Sigma A^T)^{-1} (\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b}) \end{aligned}$$

och

$$\begin{aligned} \sqrt{\det \Sigma} \cdot \det A &= \sqrt{\det \Sigma (\det A)^2} \\ &= \sqrt{\det A \det \Sigma \det A} \\ &= \sqrt{\det A \det \Sigma \det A^T} \\ &= \sqrt{\det A\Sigma A^T} \end{aligned}$$

så får vi att

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2} \sqrt{\det A\Sigma A^T}} e^{-\frac{1}{2}(\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b})^T (A\Sigma A^T)^{-1}(\mathbf{y} - A\boldsymbol{\mu} - \mathbf{b})},$$

dvs. att $\mathbf{Y} \sim N(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$.

Om $k < m$ kan vi komplettera A med en matris $\tilde{A} \in \mathbb{R}^{m-k \times m}$ så att

$$\begin{bmatrix} A \\ \tilde{A} \end{bmatrix}$$

är en kvadratisk matris av full rang. Om vi sätter

$$\mathbf{Z} = \begin{bmatrix} A \\ \tilde{A} \end{bmatrix} \mathbf{X} + \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}$$

får vi då att

$$\mathbf{Z} \sim N\left(\begin{bmatrix} A\boldsymbol{\mu} + \mathbf{b} \\ \tilde{A}\boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} A\Sigma A^T & A\Sigma \tilde{A}^T \\ \tilde{A}\Sigma A^T & \tilde{A}^T \Sigma \tilde{A} \end{bmatrix}\right).$$

Sats 1.2 ger då

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b} = \mathbf{Z}_{\{1,2,\dots,k\}} \sim N(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T),$$

vilket skulle visas. □

När g inte är injektiv på \mathcal{A} kan man fortfarande bestämma fördelningen för $U = g(\mathbf{X})$ i många fall, men det kräver en del arbete. Först måste vi dela upp \mathcal{A} i en partition

$$\mathcal{A} = A_0 \cup A_1 \cup \dots \cup A_k$$

så att $P(\mathbf{X} \in A_0) = 0$ och g bildar en bijektion mellan A_i och \mathcal{B} för varje $i = 1, 2, \dots, k$ (notera att A_0 kan vara tom). Om vi betecknar restriktionen av g till A_i som g_i för $i = 1, 2, \dots, k$ har vi då att g_i^{-1} inverterar g på A_i , dvs. att $g_i^{-1}(\mathbf{u}) \in A_i$ och $g(g_i^{-1}(\mathbf{u})) = \mathbf{u}$ för alla $\mathbf{u} \in \mathcal{B}$. Låt J_i vara jacobideterminanten för g_i^{-1} . Då har vi att

$$f_U(\mathbf{u}) = \sum_{i=1}^k f_{\mathbf{X}}(g_i^{-1}(\mathbf{u})) |J_i|$$

för all $\mathbf{u} \in \mathcal{B}$. Vi kan se detta som att för varje \mathbf{u} så lägger vi ihop täthetsvärdena för alla möjliga \mathbf{x} som kan ha gett upphov till \mathbf{u} och skalar dessa med hjälp av faktorn $|J_i|$.

Exempel 2.8. Låt X och Y vara oberoende $N(0, 1)$ -fördelade stokastiska variabler och sätt $U = X/Y$ samt $V = |Y|$ (värdet på U då $Y = 0$ kan sättas godtyckligt eftersom $P(Y = 0) = 0$). Eftersom punkterna (x, y) och $(-x, -y)$ avbildas på samma punkt $(x/y, y)$ är funktionen inte injektiv och vi måste använda formeln ovan. Vi tar $A_0 = \{(x, y) : y = 0\}$, $A_1 = \{(x, y) : y > 0\}$ och $A_2 = \{(x, y) : y < 0\}$. Eftersom $A_0 \cup A_1 \cup A_2 = \mathcal{A} = \mathbb{R}^2$ är detta en giltig partition och vi har att avbildningen är injektiv på A_1 och A_2 . Vidare har vi att restriktionerna $g_i : A_i \rightarrow \mathbb{R}^2$ är bijektiva på $\mathcal{B} = \{(u, v) : v > 0\}$ eftersom $v = |y|$ alltid är positiv och $u = x/y$ kan anta alla värden i \mathbb{R} .

Inversen g_1^{-1} ges av

$$\begin{cases} x = uv \\ y = v \end{cases}$$

medan för g_2^{-1} har vi

$$\begin{cases} x = -uv \\ y = v \end{cases}.$$

I båda fall får vi jacobideterminanten $J_1 = J_2 = v$, vilket ger

$$\begin{aligned} f_{U,V}(u, v) &= \frac{1}{2\pi} e^{-(uv)^2/2} e^{-v^2/2} |v| + \frac{1}{2\pi} e^{-(-uv)^2/2} e^{-v^2/2} |v| \\ &= \frac{v}{\pi} e^{-(u^2+1)v^2/2} \end{aligned}$$

för $(u, v) \in \mathcal{B}$ (vi kunde ersätta $|v|$ med v eftersom $v > 0$ för $(u, v) \in \mathcal{B}$).

Vi kan nu bestämma täthetsfunktionen för U genom att marginalisera bort V :

$$\begin{aligned} f_U(u) &= \int_0^{+\infty} \frac{v}{\pi} e^{-(u^2+1)v^2/2} dv \\ &= \frac{1}{2\pi} \int_0^{+\infty} e^{-(u^2+1)t/2} dt \\ &= \frac{1}{2\pi} \left[-\frac{e^{-(u^2+1)t/2}}{\frac{u^2+1}{2}} \right]_{t=0}^{+\infty} \\ &= \frac{1}{\pi} \frac{1}{u^2+1} \end{aligned}$$

där vi har använt oss av variabelbytet $t = v^2$ för att beräkna integralen. Marginalfördelningen för $U = X/Y$ som vi beräknade ovan kallas för *Cauchyfördelningen* och kommer upp i olika sammanhang. Bland annat är det en av de enklare fördelningarna som helt saknar väntevärde (den har däremot en median).

2.3 Betingning och oberoende

Vi såg i föregående föreläsning att vi kan beräkna marginalfördelningen för någon delvektor \mathbf{X}_S av \mathbf{X} där $S \subset \{1, 2, \dots, m\}$. Detta säger oss hur en eller flera av de stokastiska variablerna som utgör \mathbf{X} beter sig i isolation. En annan viktig aspekt av stokastiska vektorer är hur de olika delvektorerna relaterar till varandra. Ett naturligt sätt att beskriva dessa relationer är genom betingning (hur en viss del av den stokastiska vektorn beror på en annan del). I vissa fall beror komponenterna inte alls på varandra och vi säger i detta fall att de är oberoende.

2.3.1 Betingade fördelningar

En liknande situation kan uppstå när vi har fixerat värdena på $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m$ och sedan betraktar X_i . I båda fall får vi en fördelning för X_i , den *betingade fördelningen* för X_i givet $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m$ – det som skiljer sig är hur vi behandlar de resterande komponenterna i den stokastiska vektorn. För marginalfördelningen har vi att de resterande komponenterna kan ta vilka värden som helst men för den betingade fördelningen tillåts de endast ta en viss uppsättning värden.

Som för marginalfördelningen kan vi definiera den betingade fördelningen för en godtycklig delvektor \mathbf{X}_S för $S = \{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, m\}$. I så fall har vi den *betingade sannolikhetsfunktionen* (i det diskreta fallet) eller den *betingade täthetsfunktionen* (i det kontinuerliga fallet)

$$f_{\mathbf{X}_S|\mathbf{X}_T}(\mathbf{x}_S | \mathbf{x}_T) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_T}(\mathbf{x}_T)}$$

där $T = \{1, 2, \dots, m\} \setminus S$. Vi använder notationen $f_{\mathbf{X}_S|\mathbf{X}_T}(\mathbf{x}_S | \mathbf{x}_T)$ för att beteckna (den betingade) sannolikhetsfunktionen eller täthetsfunktionen för \mathbf{X}_S givet $\mathbf{X}_T = \mathbf{x}_T$, dvs. då de resterande komponenterna av \mathbf{X} tar värdena \mathbf{x}_T . När vi diskuterar själva fördelningen (dvs. om den är normalfördelad eller något liknande) så skriver vi detta som $\mathbf{X}_S | \mathbf{X}_T = \mathbf{x}_T$. I det tvådimensionella fallet kan vi förenkla formeln ovan till

$$f_{X_1|X_2}(x_1 | x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)},$$

vilken borde vara bekant från föregående kurs.

Exempel 2.9. Låt oss anta att vi har tre glödlampor och låt X_i vara tiden då den i :te glödlampan brinner ut. Dessa bildar då en stokastisk vektor $\mathbf{X} = [X_1, X_2, X_3]^T$. Vi antar vidare att den simultana täthetsfunktionen ges av

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{1}{4}e^{-(x_2+x_3)/2} \quad \text{för } 0 < x_1 < x_2 < +\infty \text{ och } 0 < x_1 < x_3 < +\infty.$$

Vi kan använda denna för att beräkna marginalfördelningen för (X_1, X_2) . Om $x_1 \leq 0$ har vi att $f_{X_1, X_2, X_3}(x_1, x_2, x_3) = 0$, så $f_{X_1, X_2}(x_1, x_2) = 0$ i detta fall. Om $x_1 > 0$ har vi att $f_{X_1, X_2, X_3}(x_1, x_2, x_3) > 0$ om och endast om $x_2 > x_1$ och $x_3 > x_1$. Detta ger då

$$f_{X_1, X_2}(x_1, x_2) = \int_{\mathbb{R}} f_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_3 = \int_{x_1}^{+\infty} \frac{1}{4}e^{-(x_2+x_3)/2} dx_3 = \frac{1}{4}e^{-x_2/2} \left[-2e^{-x_3/2} \right]_{x_1}^{+\infty} = \frac{1}{2}e^{-(x_1+x_2)/2}.$$

Vi har alltså den marginella täthetsfunktionen

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2}e^{-(x_1+x_2)/2} \quad \text{för } 0 < x_1 < x_2.$$

Vi kan också beräkna marginalfördelningen för X_1 antingen genom att marginalisera bort X_2 och X_3 i den stokastiska vektorn $[X_1, X_2, X_3]^T$ eller genom att marginalisera bort X_2 i $[X_1, X_2]^T$ – resultatet är detsamma. För enkelhets skull väljer vi det senare alternativet och får

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1, X_2}(x_1, x_2) dx_2 = \int_{x_1}^{+\infty} \frac{1}{2}e^{-(x_1+x_2)/2} dx_2 = \frac{1}{2}e^{-x_1/2} \left[-2e^{-x_2/2} \right]_{x_1}^{+\infty} = e^{-x_1}$$

för $x_1 > 0$. Med andra ord är X_1 exponentialfördelad med parameter ett.

Med hjälp av dessa marginalfördelningar kan vi nu beräkna fördelningen för X_2 givet $X_1 = x_1$

$$f_{X_2|X_1}(x_2 | x_1) = \frac{\frac{1}{2}e^{-(x_1+x_2)/2}}{e^{-x_1}} = \frac{1}{2}e^{-(x_2-x_1)/2}$$

för $x_2 > x_1$. Med andra ord har vi att $X_2 - x_1 | X_1 = x_1 \sim \text{Exp}(1/2)$, dvs. att för ett fixt X_1 så är tiden mellan X_1 och X_2 exponentialfördelad.

På samma sätt (genom att beräkna marginalfördelningen för (X_1, X_3) och sedan betinga med avseende på $X_1 = x_1$) får vi att $X_3 - x_1 | X_1 = x_1 \sim \text{Exp}(1/2)$. Vi kan också betinga både X_2 och X_3 med avseende på X_1 och får då

$$f_{X_2, X_3|X_1}(x_2, x_3 | x_1) = \frac{f_{X_1, X_2, X_3}(x_1, x_2, x_3)}{f_{X_1}(x_1)} = \frac{\frac{1}{4}e^{-(x_2+x_3)/2}}{e^{-x_1}} = \frac{1}{4}e^{-(x_2+x_3-2x_1)/2}$$

för $x_2, x_2 > x_1$. Om vi nu i sin tur marginaliserar denna med avseende på X_3 återfår vi den betingade fördelningen för X_2 med avseende på X_1 :

$$f_{X_2|X_1}(x_2 | x_1) = \int_{x_1}^{+\infty} \frac{1}{4} e^{-(x_2+x_3-2x_1)/2} dx_3 = \frac{1}{4} e^{-(x_2-2x_1)/2} \left[-2e^{-x_3/2} \right]_{x_1}^{+\infty} = \frac{1}{2} e^{-(x_2-x_1)/2}$$

Precis som för marginalfördelningarna kan betingade fördelningar enkelt räknas ut när det gäller delvektorer hos den flerdimensionella normalfördelningen.

Sats 2.3. Låt $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ för någon $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma \in \mathbb{R}^{m \times m}$ (positivt definit). Om $S \subset \{1, 2, \dots, m\}$ har vi

$$\mathbf{X}_S | \mathbf{X}_T = \mathbf{x}_T \sim N(\boldsymbol{\mu}_S + \Sigma_{ST} \Sigma_{TT}^{-1} \boldsymbol{\mu}_T, \Sigma_{SS} - \Sigma_{ST} \Sigma_{TT}^{-1} \Sigma_{TS})$$

där $T = \{1, 2, \dots, m\} \setminus S$, $\Sigma_{SS} = [\Sigma_{ij}]_{i,j \in S}$, $\Sigma_{ST} = [\Sigma_{ij}]_{i \in S, j \in T}$, $\Sigma_{TS} = \Sigma_{ST}^T$ och $\Sigma_{TT} = [\Sigma_{ij}]_{i,j \in T}$.

Bevis. På samma sätt som i beviset för Sats 1.2 kan vi skriva om täthetsfunktionen $f_{\mathbf{X}}(\mathbf{x})$ med hjälp av Schurkomplementet Σ/Σ_{TT} och får då

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{(d-k)/2} \sqrt{\det \Sigma_{TT}}} e^{-\frac{1}{2}(\mathbf{x}_T - \boldsymbol{\mu}_T)^T \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T)} \\ \frac{1}{(2\pi)^{k/2} \sqrt{\det(\Sigma/\Sigma_{TT})}} e^{-\frac{1}{2}(\mathbf{x}_S - \boldsymbol{\mu}_S - \Sigma_{ST} \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T))^T (\Sigma/\Sigma_{TT})^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S - \Sigma_{ST} \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T))}.$$

(Notera att vi har bytt plats på S och T .) Sats 1.2 säger nu att

$$f_{\mathbf{X}_T}(\mathbf{x}_T) = \frac{1}{(2\pi)^{(d-k)/2} \sqrt{\det \Sigma_{TT}}} e^{-\frac{1}{2}(\mathbf{x}_T - \boldsymbol{\mu}_T)^T \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T)},$$

vilket ger

$$f_{\mathbf{X}_S|\mathbf{X}_T}(\mathbf{x}_S | \mathbf{x}_T) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_T}(\mathbf{x}_T)} \\ = \frac{1}{(2\pi)^{k/2} \sqrt{\det(\Sigma/\Sigma_{TT})}} e^{-\frac{1}{2}(\mathbf{x}_S - \boldsymbol{\mu}_S - \Sigma_{ST} \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T))^T (\Sigma/\Sigma_{TT})^{-1} (\mathbf{x}_S - \boldsymbol{\mu}_S - \Sigma_{ST} \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T))}.$$

Vi har alltså att

$$\mathbf{X}_S | \mathbf{X}_T = \mathbf{x}_T \sim N(\boldsymbol{\mu}_S + \Sigma_{ST} \Sigma_{TT}^{-1}(\mathbf{x}_T - \boldsymbol{\mu}_T), \Sigma_{SS} - \Sigma_{ST} \Sigma_{TT}^{-1} \Sigma_{TS}),$$

vilket skulle visas. □

Exempel 2.10. Låt $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ där $\boldsymbol{\mu} = [1, 0, 2]^T$ och

$$\Sigma = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 2 & 3 \\ 4 & 3 & 9 \end{bmatrix}.$$

Marginalfördelningen för X_1 läses ut direkt från $\boldsymbol{\mu}$ och Σ och är alltså

$$X_1 \sim N(1, 3).$$

För den betingade fördelningen för X_1 givet $X_2 = -1, X_3 = 0$ har vi

$$\boldsymbol{\mu}_{\{1\}} + \Sigma_{\{1\},\{2,3\}} \Sigma_{\{2,3\},\{2,3\}}^{-1} ([-1, 0]^T - \boldsymbol{\mu}_{\{2,3\}}) = 1 + \begin{bmatrix} 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 9 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -2 \end{bmatrix} \\ = 1 - \frac{7}{9} = \frac{2}{9}$$

och

$$\begin{aligned}\Sigma_{\{1\},\{1\}} - \Sigma_{\{1\},\{2,3\}}\Sigma_{\{2,3\},\{2,3\}}^{-1}\Sigma_{\{2,3\},\{1\}} &= 3 - \frac{1}{9} \begin{bmatrix} 1 & 4 \end{bmatrix} \begin{bmatrix} 9 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \end{bmatrix}^T \\ &= 3 - \frac{17}{9} = \frac{10}{9}.\end{aligned}$$

Vi får alltså

$$X_1 \mid X_2 = -1, X_3 = 0 \sim N(2/9, 10/9).$$

Vi kan notera två saker här. Den första är att väntevärdet för den betingade fördelningen är mindre än för marginalfördelningen. Detta kan förklaras med att både X_2 och X_3 är starkt korrelerade med X_1 och eftersom både X_2 och X_3 har värden under deras respektive väntevärden så drar de ner väntevärdet på X_1 . Vi kan också notera att variansen är mindre ($10/9$ jämfört med 3). Detta beror också på att vi har fixerat X_2 och X_3 , vilket reducerar den totala mängden varians i den stokastiska vektorn.

Om vi vänder på definitionen av den betingade fördelningen får vi ett användbart redskap

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_S|\mathbf{X}_T}(\mathbf{x}_S \mid \mathbf{x}_T)f_{\mathbf{X}_T}(\mathbf{x}_T).$$

Denna relation kallas för *kedjeregeln* (ej att förväxla med kedjeregeln inom differentialkalkylen) och låter oss förenkla vissa uttryck och bygga modeller.

Exempel 2.11. Låt oss återta exemplet med glödlamporna. Vi ansätter att brinntiden för lampa ett har fördelningen $\text{Exp}(1)$, dvs.

$$f_{X_1}(x_1) = e^{-x_1}$$

för $x_1 > 0$. Lampa två antar vi har brinntid med fördelning $\text{Exp}(1/2)$. Om vi betecknar denna med Y_2 får vi

$$f_{Y_2}(y_2) = \frac{1}{2}e^{-y_2/2}$$

för $y_2 > 0$.

Men vi sätter ju inte på lampa två förrän lampa ett släcks. Om detta händer vi tidpunkt x_1 så har vi att lampa två släcks vid tiden $Y_2 + x_1$, vilket har fördelningen

$$f_{Y_2+x_1}(x_2) = \frac{1}{2}e^{-(x_2-x_1)/2}$$

för $x_2 > x_1$. (Detta fås genom enkel transformation av en endimensionell variabel.) Men detta är ju precis den betingade fördelningen för X_2 (tiden tills lampa två släcks) givet $X_1 = x_1$, dvs.

$$f_{X_2|X_1}(x_2 \mid x_1) = \frac{1}{2}e^{-(x_2-x_1)/2}$$

för $x_2 > x_1$.

Kedjeregeln ger nu att

$$f_{X_1,X_2}(x_1, x_2) = f_{X_2|X_1}(x_2 \mid x_1)f_{X_1}(x_1) = \frac{1}{2}e^{-(x_2-x_1)/2} \cdot e^{-x_1} = \frac{1}{2}e^{-(x_1+x_2)/2}$$

för $x_2 > x_1$, vilket stämmer överens med marginalfördelningen vi fick för (X_1, X_2) tidigare.

2.3.2 Oberoende stokastiska variabler

Ett viktigt specialfall av betingade fördelningar uppstår då den betingade fördelningen sammanfaller med marginalfördelningen. Vi har alltså, för $S \subset \{1, 2, \dots, m\}$ och $T = \{1, 2, \dots, m\} \setminus S$, att

$$f_{\mathbf{X}_S|\mathbf{X}_T}(\mathbf{x}_S \mid \mathbf{x}_T) = f_{\mathbf{X}_S}(\mathbf{x}_S).$$

Med andra ord så ändras inte fördelningen av \mathbf{X}_S om vi låter \mathbf{X}_T vara fixt eller variera. Genom definitionen av den betingade fördelningen har vi

$$\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_T}(\mathbf{x}_T)} = f_{\mathbf{X}_S}(\mathbf{x}_S) \quad \Leftrightarrow \quad f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_S}(\mathbf{x}_S)f_{\mathbf{X}_T}(\mathbf{x}_T).$$

När detta gäller säger vi att \mathbf{X}_S och \mathbf{X}_T är *oberoende* stokastiska vektorer och betecknar detta med $\mathbf{X}_S \perp \mathbf{X}_T$.

Notera att definitionen ovan kan också tillämpas då $S \cup T \neq \{1, 2, \dots, m\}$. I detta fall betraktas oberoendet mellan \mathbf{X}_S och \mathbf{X}_T med avseende på $\mathbf{X}_{S \cup T}$ och vi har

$$f_{\mathbf{X}_{S \cup T}}(\mathbf{x}_{S \cup T}) = f_{\mathbf{X}_S}(\mathbf{x}_S) f_{\mathbf{X}_T}(\mathbf{x}_T).$$

Ett viktigt specialfall är då $S = \{i\}$ och $T = \{j\}$ och vi får

$$f_{X_i, X_j}(x_i, x_j) = f_{X_i}(x_i) f_{X_j}(x_j),$$

vilket är definitionen av oberoende för endimensionella stokastiska variabler från föregående kurs.

Om vi har fler än två delvektorer $\mathbf{X}_{S_1}, \mathbf{X}_{S_2}, \dots, \mathbf{X}_{S_k}$ säger vi att dessa är oberoende om alla par av disjunkta unioner är oberoende, dvs. att vi har

$$\mathbf{X}_{\cup_{i \in P} S_i} \perp \mathbf{X}_{\cup_{j \in Q} S_j}$$

för alla $P, Q \subset \{1, 2, \dots, k\}$ så att $P \cap Q = \emptyset$. (Notera att engelskan skiljer på denna definition, "mutual independence", och fallet då endast par av delvektorer är oberoende, "pairwise independent". På svenska syftar oberoende alltid på den tidigare definitionen.)

2.4 Väntevärde, varians och kovarians

Precis som i det endimensionella fallet är det användbart att försöka sammanfatta flerdimensionella stokastiska variabler. Detta görs enklast med väntevärden som beskriver vilka värden den stokastiska vektorn tar "i genomsnitt". För att beskriva spridningen blir saken mer komplicerad än för endimensionella stokastiska variabler då vi har flera möjliga riktningar som vektorn kan sprida sig i. Vi behöver därför en matris för att beskriva denna.

2.4.1 Väntevärde i flera dimensioner

Precis som i det endimensionella fallet kan vi definiera väntevärden för flerdimensionella stokastiska variabler. Om \mathbf{X} är en diskret m -dimensionell stokastisk variabel och vi har $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$ definierar vi väntevärdet

$$\mathbb{E}[g(\mathbf{X})] = \sum_{x_1, x_2, \dots, x_m} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}).$$

I det kontinuerliga fallet har vi

$$\mathbb{E}[g(\mathbf{X})] = \int_{\mathbb{R}^m} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Om vi bara har $g(\mathbf{x}) = \mathbf{x}$ talar vi om väntevärdet $\mathbb{E}[\mathbf{X}]$ av \mathbf{X} . Väntevärdet är linjär i $g(\mathbf{X})$, dvs.

$$\mathbb{E}[A g(\mathbf{X}) + \mathbf{b}] = A \mathbb{E}[g(\mathbf{X})] + \mathbf{b},$$

för $A \in \mathbb{R}^{k \times p}$ och $\mathbf{b} \in \mathbb{R}^k$.

Definitionen ovan kan också generaliseras till en matrisvärd funktion $g : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times q}$ genom samma summa- och integralformler ovan. Vi har då den ytterligare linjäritetsegenskapen

$$\mathbb{E}[A g(\mathbf{X}) B + C] = A \mathbb{E}[g(\mathbf{X})] B + C$$

för matriser $A \in \mathbb{R}^{k \times p}$, $B \in \mathbb{R}^{q \times s}$ och $C \in \mathbb{R}^{k \times s}$.

För multinomialfördelningen har vi att om $\mathbf{X} \sim \text{Multi}(n, p_1, p_2, \dots, p_m)$ så gäller

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_m \end{bmatrix}$$

(se övning 2).

Exempel 2.12. Låt $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$ vara den m -dimensionella standardnormalfördelningen. I detta fall har vi

$$\begin{aligned} E[\mathbf{Z}] &= \int_{\mathbb{R}^m} \mathbf{z} \cdot \frac{1}{(2\pi)^{m/2}} e^{-\|\mathbf{z}\|^2/2} d\mathbf{z} \\ &= \frac{1}{(2\pi)^{m/2}} \int_{\mathbb{R}^m} \mathbf{z} e^{-\sum_{i=1}^m z_i^2/2} d\mathbf{z}. \end{aligned}$$

Det j :te elementet av $E[\mathbf{Z}]$ ges då av

$$\begin{aligned} &\frac{1}{(2\pi)^{m/2}} \int_{\mathbb{R}^m} z_j \prod_{i=1}^m e^{-z_i^2/2} d\mathbf{z} \\ &= \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} z_j e^{-z_j^2/2} dz_j \cdot \frac{1}{(2\pi)^{(m-1)/2}} \int_{\mathbb{R}^{m-1}} \prod_{i=1, i \neq j}^m e^{-z_i^2/2} dz_1 dz_2 \cdots dz_{j-1} dz_{j+1} \cdots dz_m \\ &= 0 \cdot 1 = 0. \end{aligned}$$

Med andra ord har vi $E[\mathbf{Z}] = \mathbf{0}$.

Man kan använda detta resultat tillsammans med Sats 2.1 för att visa att normalfördelningens $N(\boldsymbol{\mu}, \Sigma)$ väntevärde är $\boldsymbol{\mu}$ (föga oförväntat, se övning 2).

2.4.2 Kovariansmatris

Som vi nämnde ovan så kan spridningen hos en stokastisk vektor inte sammanfattas med ett värde eller ens en vektor – vad som krävs är en matris som beskriver hur fördelningen är utspridd i olika riktningar. Vi kallar denna matris *kovariansmatrisen* och definierar den som

$$\text{Cov}[\mathbf{X}] = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T].$$

Andra namn på matrisen är *autokovariansmatris*, *variansmatris* och *varians-kovariansmatris* och den betecknas ibland $\text{Var}[\mathbf{X}]$.

Det första vi ser är att $\text{Cov}[\mathbf{X}]$ är en $m \times m$ -matris. Vidare ser vi att $\text{Cov}[\mathbf{X}]^T = \text{Cov}[\mathbf{X}]$ är symmetrisk. Vi ser också att det (i, j) :te elementet av $\text{Cov}[\mathbf{X}]$ ges av

$$E[(X_i - E[X_i])(X_j - E[X_j])] = \text{Cov}[X_i, X_j],$$

därför namnet kovariansmatris. På diagonalen har vi varianserna $\text{Cov}[X_i, X_i] = \text{Var}[X_i]$.

En annan viktig egenskap är att $\text{Cov}[\mathbf{X}]$ är positivt semidefinit. Vi ser detta genom att ta

$$\begin{aligned} \mathbf{u}^T \text{Cov}[\mathbf{X}] \mathbf{u} &= E[\mathbf{u}^T (\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T \mathbf{u}] \\ &= E[|\mathbf{u}^T (\mathbf{X} - E[\mathbf{X}])|^2] \geq 0 \end{aligned}$$

för godtycklig $\mathbf{u} \in \mathbb{R}^m$. Med andra ord är alla egenvärden icke-negativa (men inte nödvändigtvis positiva).

Kovariansmatriser är inte linjära (som väntevärdet), men uppfyller en liknande egenskap. Vi har nämligen att för $A \in \mathbb{R}^{k \times m}$ och $\mathbf{b} \in \mathbb{R}^k$ så gäller

$$\text{Cov}[A\mathbf{X} + \mathbf{b}] = E[(A\mathbf{X} - A E[\mathbf{X}])(A\mathbf{X} - A E[\mathbf{X}])^T]$$

eftersom $E[A\mathbf{X} + \mathbf{b}] = A E[\mathbf{X}] + \mathbf{b}$. Om vi vidareutvecklar får vi

$$\text{Cov}[A\mathbf{X} + \mathbf{b}] = E[A(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T A^T] = A \text{Cov}[\mathbf{X}] A^T.$$

Detta generaliserar den endimensionella relationen $\text{Var}[aX + b] = a^2 \text{Var}[X]$.

Ibland kan vi förenkla beräkningen av en kovariansmatris genom följande omskrivning

$$\begin{aligned} \text{Cov}[\mathbf{X}] &= E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] \\ &= E[\mathbf{X}\mathbf{X}^T] - E[\mathbf{X}] E[\mathbf{X}]^T - E[\mathbf{X}] E[\mathbf{X}^T] + E[\mathbf{X}] E[\mathbf{X}]^T \\ &= E[\mathbf{X}\mathbf{X}^T] - E[\mathbf{X}] E[\mathbf{X}]^T. \end{aligned}$$

För multinomialfördelningen har vi att om $\mathbf{X} \sim \text{Multi}(n, p_1, p_2, \dots, p_m)$ får vi (se övning 2)

$$\text{Cov}[\mathbf{X}] = \begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_m \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_m \\ \vdots & \vdots & \ddots & \vdots \\ -np_mp_1 & -np_mp_2 & \cdots & np_m(1-p_m) \end{bmatrix}.$$

Exempel 2.13. Låt $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$ vara den m -dimensionella standardnormalfördelningen (igen). Vi vet att det (i, j) :te elementet av $\text{Cov}[\mathbf{Z}]$ ges av $\text{Cov}[Z_i, Z_j]$. Eftersom $(Z_i, Z_j) \sim N(\mathbf{0}, \mathbf{I})$ enligt Sats 1.2 har vi för $i \neq j$

$$\text{Cov}[Z_i, Z_j] = \int_{\mathbb{R}^2} z_i z_j \cdot \frac{1}{2\pi} e^{-(z_i^2 + z_j^2)/2} dz_i dz_j = \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z_i e^{-z_i^2/2} dz_i \right) \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z_j e^{-z_j^2/2} dz_j \right) = 0 \cdot 0 = 0.$$

På samma sätt har vi att $\text{Cov}[Z_i, Z_i] = \text{Var}[Z_i] = 1$. Detta ger kovariansmatrisen

$$\text{Cov}[\mathbf{Z}] = \mathbf{I}.$$

Återigen kan vi använda detta resultat tillsammans med Sats 2.1 för att visa att kovariansmatrisen för $N(\boldsymbol{\mu}, \Sigma)$ är lika med Σ .

För två delvektorer \mathbf{X}_S och \mathbf{X}_T kan vi också definiera kovariansmatrisen

$$\text{Cov}[\mathbf{X}_S, \mathbf{X}_T] = E[(\mathbf{X}_S - E[\mathbf{X}_S])(\mathbf{X}_T - E[\mathbf{X}_T])^T]$$

mellan dessa vektorer. Detta motsvarar så klart blocket av $\text{Cov}[\mathbf{X}]$ där raderna tas från S och kolonnerna tas från T .

Om dessa två delvektorer är obereonde (dvs. $\mathbf{X}_S \perp \mathbf{X}_T$) har vi att

$$\text{Cov}[\mathbf{X}_S, \mathbf{X}_T] = E[(\mathbf{X}_S - E[\mathbf{X}_S])(\mathbf{X}_T - E[\mathbf{X}_T])^T] = E[\mathbf{X}_S - E[\mathbf{X}_S]] E[\mathbf{X}_T - E[\mathbf{X}_T]]^T = \mathbf{0} \cdot \mathbf{0}^T = \mathbf{0}$$

eftersom alla element av \mathbf{X}_S är oberoende av alla element i \mathbf{X}_T och vi då kan faktorisera väntevärdet.

Konversen till detta resultat håller inte i allmänhet, men det finns ett viktigt specialfall.

Sats 2.4. Låt $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ för något $\boldsymbol{\mu} \in \mathbb{R}^m$ och $\Sigma \in \mathbb{R}^{m \times m}$ (positivt definit). Om $S, T \subset \{1, 2, \dots, m\}$ så att $S \cap T = \emptyset$ så har vi att

$$\mathbf{X}_T \perp \mathbf{X}_S \Leftrightarrow \text{Cov}[\mathbf{X}_S, \mathbf{X}_T] = \mathbf{0}.$$

Bevis. Den ena riktningen har vi redan visat. Den andra fås genom att faktorisera den simultana täthetsfunktionen och visa att den kan fås som produkten av täthetsfunktionerna för \mathbf{X}_S och \mathbf{X}_T . \square