Föreläsning 07

Statistisk inlärning och dataanalys (Kungliga Tekniska Högskolan)

Föreläsning 7

7.1 Konjugerade apriorifördelningar

I föregående föreläsning repeterades ett mönster för de olika datafördelningarna vi studerade. För varje datafördelning valde vi en apriorifördelning för parametern $\boldsymbol{\Theta}$ från en viss familj (betafördelningar, gammafördelningar, osv.) och det visade sig sedan att aposteriorifördelningen sedan kom från samma familj. När detta händer säger vi att denna familj av apriorifördelningar är *konjugerad* med avseende på datafördelningens familj.

Vi har då sett följande exempel

- $\mathrm{Bin}(n, \theta)$ har som konjugerade apriorifördelningar $\mathrm{Beta}(\alpha, \beta)$,
- $\mathrm{Po}(\theta)$ har som konjugerade apriorifördelningar $\mathrm{Gamma}(\alpha, \lambda)$,
- $\mathrm{Exp}(\theta)$ har som konjugerade apriorifördelningar $\mathrm{Gamma}(\alpha, \lambda)$, och
- $\mathrm{N}(\theta, \sigma^2)$ har som konjugerade apriorifördelningar $\mathrm{N}(\mu, \tau^2)$.

Notera att enligt denna definition finns det en familj som är konjugerad med avseende på alla datafördelningar: familjen av alla fördelningar över parameterrummet Ω (eftersom aposteriorifördelningen nödvändigtvis ingår i denna familj). För att undvika detta används ofta begreppet *naturlig konjugerad apriorifördelning*. Med detta menas att täthetsfunktionen (eller sannolikhetsfunktionen) hos apriorifördelningen har samma funktionella form (dvs. samma uttryck i θ upp till en normaliseringskonstant) som täthetsfunktionen (eller sannolikhetsfunktionen) hos datafördelningen. Alla exempel vi gav ovan är naturliga konjugerade apriorifördelningar.

Exempel 7.1. Låt oss undersöka en annan datafördelning: $X \mid \Theta = \theta \sim \mathrm{N}(\mu, \theta)$, dvs. normalfördelad data med känt väntevärde μ och där variansen θ är vår parameter. I detta fall har vi

$$f_{X|\Theta}(x \mid \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\mu)^2}.$$

Om vi bortser från alla konstanter (dvs. faktorer some inte beror på θ) får vi

$$f_{X|\Theta}(x \mid \theta) = a\theta^{-b} e^{-c/\theta}.$$

När vi ser över de olika fördelningar vi har i formelsamlingen ser vi att detta överenstämmer med den inversa gammafördelningen som har täthetsfunktion

$$f_\Theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha} \theta^{-\alpha-1} e^{-\beta/\theta}.$$

Om vi nu sätter $\Theta \sim \mathrm{InvGamma}(\alpha_0, \beta_0)$ så får vi aposteriorifördelningen

$$\begin{aligned} f_{\Theta|X}(\theta \mid x) &\propto \theta^{-1/2} e^{-\frac{1}{2\theta}(x-\mu)^2} \cdot \theta^{-\alpha_0-1} e^{-\beta_0/\theta} \\ &= \theta^{-(\alpha_0+\frac{1}{2})-1} e^{-(\beta_0+\frac{1}{2}(x-\mu)^2)/\theta}. \end{aligned}$$

Med andra ord har vi att $\Theta \mid X = x \sim \mathrm{InvGamma}(\alpha_1, \beta_1)$ där $\alpha_1 = \alpha_0 + 1/2$ och $\beta_1 = \beta_0 + (x-\mu)^2/2$.

Detta resultat generaliseras till godtyckligt antal data $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}}$ för betingat oberoende $X_1, X_2, \ldots, X_n$ givet Θ där $X_i \mid \Theta = \theta \sim \mathrm{N}(\mu, \theta)$ för $i = 1, 2, \ldots, n$ där $\Theta \sim \mathrm{InvGamma}(\alpha_0, \beta_0)$. Vi får då att

$$\Theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{InvGamma}(\alpha_n, \beta_n)$$

där

$$\alpha_n = \alpha_0 + \frac{n}{2}$$
$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Notera att naturliga konjugerade apriorifördelningar inte alltid finns. För de flesta datafördelningarna som vi studerar i denna kurs kommer de dock finnas.

Ett större problem med konjugerade apriorifördelningar är att de är relativt oflexibla. Låt oss säga att vi har binomialfördelad data, dvs. att $X \mid \Theta = \theta \sim \mathrm{Bin}(n, \theta)$ och vi vet att θ antingen är runt $1/4$ eller runt $3/4$, men inte däremellan. Den konjugerade apriorifördelningen för binomialfördelad data är betafördelningen, men inget val av parametrar ger oss en fördelning som är stor runt $1/4$ och $3/4$, men liten för andra värden på θ. En sådan bimodal fördelning är nödvändigtvis okonjugerad. I detta fall kan vi beräkna aposteriorifördelningen, men den har inte samma enkla form som för en konjugerad apriorifördelning.

7.2 Icke-informativa apriorifördelningar

Valet av apriorifördelning kan vara svårt för vissa tillämpningar. Ofta har vi ingen bra idé av hur parametern borde vara fördelad. En lösning är då att använda en apriorifördelning som innehåller så lite information som möjligt. Denna typ av apriorifördelningar kallas för *icke-informativa apriorifördelningar*.

Notera att även om målet med dessa apriorifördelningar är att påverka skattningen så lite som möjligt utgör valet av en icke-informativ apriorifördelning en viss påverkan i sig. Andra val, som hur datafördelningen är parametriserad, påverkar också aposteriorifördelningen. Exempelvis kan binomialfördelningen $\mathrm{Bin}(n, p)$ parametriseras genom standardparametern p eller genom log-oddsen $\log(p/(1-p))$ (den naturliga parametern i exponentialfamiljen). En icke-informativ fördelning över en parametrisering kan ge olika resultat i en annan parametrisering.

Standardförfarandet för att ta fram icke-informativa apriorifördelningar är att studera aposteriorifördelningen för en konjugerad apriorifördelning och låta antalet observationer gå mot oändligheten för att se för vilka parametrar som datan dominerar aposteriorifördelningen. För binomialfördelad data har vi, till exempel att $X \mid \Theta = \theta \sim \mathrm{Bin}(n, \theta)$ och $\Theta \sim \mathrm{Beta}(\alpha, \beta)$ ger

$$\Theta \mid X = x \sim \mathrm{Beta}(\alpha + x, \beta + n - x).$$

Om vi låter $x, n \to +\infty$ så får vi, approximativt, aposteriorifördelningen $\mathrm{Beta}(x, n-x)$. Med andra ord motsvarar detta $\alpha = 0$ och $\beta = 0$ i apriorifördelningen. En rimlig icke-informativ apriorifördelning är då $\mathrm{Beta}(0, 0)$.

Ett problem med denna fördelning är att den faktiskt inte är en fördelning. När vi tar $\alpha, \beta \to 0$ så divergerar täthetsfunktionen och approximerar en Bernoullifördelning med parameter $1/2$, dvs. med sannolikhet $1/2$ på 0 och 1. Med andra ord har vi att

$$f_\Theta(\theta) \propto \theta^{-1}(1-\theta)^{-1}$$

inte går att integrera och därför inte kan bilda en täthetsfunktion. En sådan ”fördelning” kallas för en *oegentlig fördelning* och när den används som apriorifördelning kalles den för en *oegentlig apriorifördelning*.

Även om en apriorifördelning är oegentlig betyder det inte att den är oanvändbar. I fallet ovan med $\mathrm{Beta}(0, 0)$ kan vi ta fram aposteriorifördelningen givet $X = x$ där $X \mid \Theta = \theta \sim \mathrm{Bin}(n, \theta)$, vilken ges av

$$f_{\Theta|X}(\theta \mid x) \propto \theta^x(1-\theta)^{n-x} \cdot \theta^{-1}(1-\theta)^{-1} = \theta^{x-1}(1-\theta)^{n-x-1},$$

dvs. $\Theta \mid X = x \sim \mathrm{Beta}(x, n-x)$. Såvida vi inte har $x = 0$ eller $x = n$ så är aposteriorifördelningen alltså en egentlig fördelning. Många andra oegentliga apriorifördelningar har samma egenskap, dvs. att aposteriorifördelningen är en egentlig fördelning för de flesta datapunkter.

Apriorifördelningen $\mathrm{Beta}(0, 0)$ kallas för *Haldanes apriorifördelning*. Andra alternativ är $\mathrm{Beta}(1, 1)$, som ger samma sannolikhet åt alla värden på θ. Denna apriorifördelning är en egentlig fördelning och kallas för *Bayes* eller *Laplaces apriorifördelning*. Ett annat alternativ är *Jeffreys apriorifördelning* som ges av $\mathrm{Beta}(1/2, 1/2)$ (också kallad för *arcus-sinusfördelningen*).

För normalfördelningen har vi att om $X_1, X_2, \ldots, X_n$ är betingat oberoende givet $\Theta = \theta$ med $X_i \mid \Theta = \theta \sim \mathrm{N}(\theta, \sigma^2)$ för $\Theta \sim \mathrm{N}(\mu_0, \tau_0^2)$ så har vi sedan tidigare att

$$\Theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{N}(\mu_n, \tau_n^2)$$

där

$$\mu_n = \left(\frac{\mu_0}{\tau_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2}\right) / \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)$$
$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

För n stort har vi att

$$\mu_n \approx \frac{\sum_{i=1}^n x_i}{n}$$
$$\tau_n^2 \approx \frac{\sigma^2}{n}.$$

Detta motsvarar att vi låter $\mu_0 \to 0$ och $\tau_0^2 \to +\infty$ i vår apriorifördelning. En rimlig icke-informativ ”apriorifördelning” är då $\mathrm{N}(0, +\infty)$, dvs. en konstant sannolikhet över hela $\mathbb{R}$:

$$f_\Theta(\theta) \propto 1.$$

Återigen är detta en oegentlig apriorifördelning, men den ger alltid en egentlig aposteriorifördelning när vi uppdaterar med data eftersom vi då får

$$\Theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{N}\left(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n}\sigma^2\right).$$

På samma sätt kan vi ta fram en icke-informativ apriorifördelning för normalfördelad data med känt väntevärde och okänd varians. Här har vi $X_1, X_2, \ldots, X_n$ betingat oberoende givet $\Theta = \theta$ och $X_i \mid \Theta = \theta \sim \mathrm{N}(\mu, \theta)$ för $i = 1, 2, \ldots, n$ och $\Theta \sim \mathrm{InvGamma}(\alpha, \beta)$, vilket ger

$$\Theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{InvGamma}\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}\right).$$

När n blir stort approximerar denna fördelning

$$\mathrm{InvGamma}(\left(\frac{n}{2}, \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}\right),$$

vilket gör att en icke-informativ apriorifördelning skulle kunna vara $\mathrm{InvGamma}(0, 0)$. Här har vi då

$$f_\Theta(\theta) \propto \frac{1}{\theta}$$

vilket ger en oegentlig fördelning, men denna ger också upphov till en egentlig aposteriorifördelning

$$\Theta \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{InvGamma}\left(\frac{n}{2}, \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}\right).$$

7.3 Flerdimensionella parametrar

Hittills har vi arbetat med en enda parameter för vilken vi har tagit fram en aposteriorifördelning. Som vi har sett tidigare i kursen är det dock ofta relevant att modellera data med hjälp av flera parametrar. Inom bayesiansk inferens innebär detta helt enkelt att vi har en flerdimensionell apriorifördelning $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ över $\Omega \subset \mathbb{R}^d$ och en datafördelning $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta})$ som ger oss en aposteriorifördelning $f_{\boldsymbol{\Theta}|\boldsymbol{X}}(\boldsymbol{\theta} \mid \boldsymbol{x})$ från vilken vi kan dra olika slutsatser om vår data. Alla resultat som vi tidigare visat för endimensionella parametrar Θ kan enkelt generaliseras för flerdimensionella parametrar $\boldsymbol{\Theta}$.

En styrka med det bayesianska synsättet är när vi väl har aposteriorifördelningen kan vi manipulera denna för att extrahera den information som är relevant för oss. Till exempel, vi kanske har två parametrar $\boldsymbol{\Theta}_1$ och $\boldsymbol{\Theta}_2$ i vår modell, men endast $\boldsymbol{\Theta}_1$ är av intresse för vår tillämpning. Parametern $\boldsymbol{\Theta}_2$ kallas då för en *störparameter*. I detta fall marginaliserar vi helt enkelt med avseende på $\boldsymbol{\Theta}_2$ för att erhålla den marginella aposteriorifördelningen för $\boldsymbol{\Theta}_1$ givet datat. På samma sätt kan vi också bilda betingade fördelningar för att studera hur vissa parametrar beror på andra.

7.3.1 Multinomialfördelad data

Som vi har sett innan är en multinomialfördelningen en användbar diskret flerdimensionell fördelning som används för att beskriva antalet gånger en viss händelse inträffar. Parametrarna är det totala antalet försök m samt sannolikheterna för de olika händelserna. Låt oss anta att vi har

$$\boldsymbol{X} \mid \boldsymbol{\Theta} = \boldsymbol{\theta} \sim \mathrm{Multi}(m, \boldsymbol{\theta})$$

där $\boldsymbol{\theta} \in \mathbb{R}^d$ så att $\theta_1 + \ldots + \theta_d = 1$. Vi har då

$$f_{\boldsymbol{X}\mid\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \frac{m!}{x_1! \cdots x_d!} \theta_1^{x_1} \cdots \theta_d^{x_d}.$$

Om vi söker en naturligt konjugerad apriorifördelning borde den vara på formen

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \propto \theta_1^{a_1} \cdots \theta_d^{a_d}$$

för $\theta_1 + \ldots + \theta_d = 1$ och $0 \leq \theta_1, \ldots, \theta_d \leq 1$. En naturlig kandidat är då Dirichletfördelningen, vars täthetsfunktion ges av

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d} \theta_1^{\alpha_1 - 1} \cdots \theta_d^{\alpha_d - 1}$$

för $\theta_1 + \ldots + \theta_d = 1$ och $0 \leq \theta_1, \ldots, \theta_d \leq 1$. Detta skrivs som $\boldsymbol{\Theta} \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_d) = \mathrm{Dirichlet}(\boldsymbol{\alpha})$. Aposteriorifördelningen ges då av

$$\boldsymbol{\Theta} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{x}).$$

Två rimliga icke-informativa apriorifördelningar är $\mathrm{Dirichlet}(0, \ldots, 0)$ (en oegentlig apriorifördelning) och $\mathrm{Dirichlet}(1, \ldots, 1)$ (den likformiga fördelningen). I båda fall kan vi använda uppdateringsregeln ovan. Observera dock att aposteriorifördelningen för $\mathrm{Dirichlet}(0, \ldots, 0)$ endast är en egentlig fördelning om $x_1, \ldots, x_d > 0$.

Exempel 7.2. En opinionsundersökning gjordes under oktober 1998 av den amerikanska TV-kanalen CBS. Svarspersonerna tillfrågades om vilken presidentkandidat de tänkte rösta på i höstens val: George H.W. Bush, Michael Dukakis eller någon annan (detta inkluderade också ingen preferens). Totalt 1447 personer deltog och resultaten var följande:

Bush	Dukakis	Annan/ingen
727	583	137

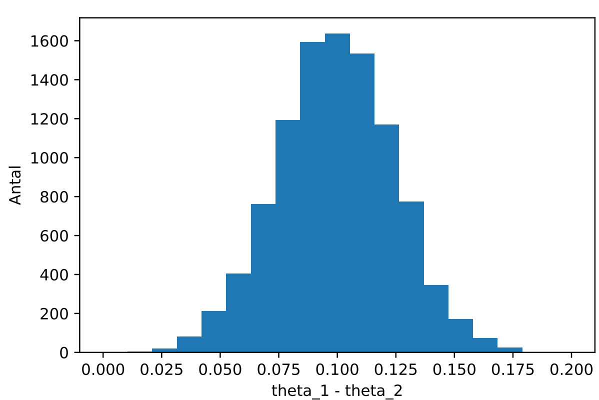
Om vi antar att svarspersonerna valdes slumpmässigt ur populationen av röstberättigade får vi att antalet svar borde vara multinomialfördelat med parameter motsvarande de andelar av population som har den underliggande preferensen. Med andra ord har vi $\boldsymbol{X} \mid \boldsymbol{\Theta} = \boldsymbol{\theta} \sim \mathrm{Multi}(1447, \boldsymbol{\theta})$.

Om vi inte har någon tidigare information om de olika kandidaternas stöd kan vi använda en icke-informativ apriorifördelning. Här tar vi $\boldsymbol{\Theta} \sim \mathrm{Dirichlet}(1, 1, 1)$, vilket ger aposteriorifördelningen

$$\boldsymbol{\Theta} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{Dirichlet}(728, 584, 138).$$

Här kan det vara intressant att beräkna fördelningen för $\Theta_1 - \Theta_2$, dvs. skillnaden i andel för Bush jämfört med Dukakis. Vi kan försöka beräkna denna fördelning analytiskt, men det är enklare att simulera utfall från Θ_1 och Θ_2, bilda differensen, och plotta dess histogram.

Om vi drar $n_{\mathrm{samp}} = 10000$ utfall får vi följande histogram:



Alla utfall har att $\theta_1 > \theta_2$, så vi kan med säkerhet säga att Bush har en klar fördel över Dukakis här.

Notera att om vi hade en annan apriorifördelning (t.ex. som gav Dukakis ett större stöd till och börja med) hade aposteriorifördelningen sett annorlunda ut.

7.3.2 Normalfördelad data

En annan viktig modell är normalfördelad data. Vi har redan studerat denna, men där vi har tagit fixt väntevärde μ eller varians σ^2. Låt oss nu undersöka fallet då både väntevärdet och variansen är okända.

Låt oss anta att $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}}$ för betingat oberoende $X_1, X_2, \ldots, X_n$ där

$$X_i \mid \Theta_{\mathrm{M}} = \theta_{\mathrm{M}}, \Theta_{\mathrm{V}} = \theta_{\mathrm{V}} \sim \mathrm{N}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}})$$

för $i = 1, 2, \ldots, n$. Om vi studerar den betingade täthetsfunktionen för $\boldsymbol{X}$ har vi

$$\begin{aligned} f_{\boldsymbol{X}\mid\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\boldsymbol{x} \mid \theta_{\mathrm{M}}, \theta_{\mathrm{V}}) &= \prod_{i=1}^{n} \frac{1}{(2\pi\theta_{\mathrm{V}})^{1/2}} e^{-(x_i-\theta_{\mathrm{M}})^2/2\theta_{\mathrm{V}}} \\ &\propto \theta_{\mathrm{V}}^{-n/2} e^{-\frac{1}{2\theta_{\mathrm{V}}}\sum_{i=1}^{n}(x_i-\theta_{\mathrm{M}})^2} \\ &= \theta_{\mathrm{V}}^{-n/2} e^{-\frac{1}{2\theta_{\mathrm{V}}}((n-1)s^2+n(\bar{x}-\theta_{\mathrm{M}})^2}, \end{aligned}$$

där vi i sista raden har använt identiteten

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - a)^2 = (n-1)s^2 + n(\bar{x} - a)^2$$

för $a \in \mathbb{R}$. För att hitta en konjugerad apriorifördelning söker vi alltså en täthetsfunktion på formen

$$f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}}) \propto \theta_{\mathrm{V}}^{-a} e^{-\frac{1}{\theta_{\mathrm{V}}}(b+c(\theta_{\mathrm{M}}-d)^2}.$$

Denna ganska komplicerade form ger inte någon självklar fördelning, men en faktor ser bekant ut:

$$\theta_{\mathrm{V}}^{-a} e^{-\frac{b}{\theta_{\mathrm{V}}}}.$$

Denna form delas av täthetsfunktionen hos den inversa gammafördelningen, som vi såg innan som den konjugerade apriorifördelningen för normalfördelad data med känt väntevärde. Om vi då tar $\Theta_{\mathrm{V}} \sim \mathrm{InvGamma}(\alpha_0, \beta_0)$ så ser vi kan bygga upp den sökta täthetsfunktionen genom att ta $\Theta_{\mathrm{M}} \mid \Theta_{\mathrm{V}} = \theta_{\mathrm{V}} \sim \mathrm{N}(\mu_0, \theta_{\mathrm{V}}/\lambda_0)$, och använda kedjeregeln för att få

$$\begin{aligned} f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}}) &= \frac{\sqrt{\lambda_0}}{(2\pi\theta_{\mathrm{V}})^{1/2}} e^{-\frac{\lambda_0}{2\theta_{\mathrm{V}}}(\theta_{\mathrm{M}}-\mu_0)^2} \cdot \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta_{\mathrm{V}}^{-\alpha_0-1} e^{-\beta_0/\theta_{\mathrm{V}}} \\ &= \frac{\sqrt{\lambda_0}\beta_0^{\alpha_0}}{(2\pi\theta_{\mathrm{V}})^{1/2}\Gamma(\alpha_0)} \theta_{\mathrm{V}}^{-\alpha_0-1} e^{-\frac{1}{2\theta_{\mathrm{V}}}(2\beta_0+\lambda_0(\theta_{\mathrm{M}}-\mu_0)^2)}. \end{aligned}$$

Denna fördelning kallas den normal- och inversa gammafördelningen och betecknas $\mathrm{NInvGamma}(\mu_0, \lambda_0, \alpha_0, \beta_0)$.

Genom att multiplicera $f_{\boldsymbol{X}\mid\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\boldsymbol{x} \mid \theta_{\mathrm{M}}, \theta_{\mathrm{V}})$ och $f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}})$ får vi

$$\begin{aligned} f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}\mid\boldsymbol{X}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}} \mid \boldsymbol{x}) &\propto f_{\boldsymbol{X}\mid\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\boldsymbol{x} \mid \theta_{\mathrm{M}}, \theta_{\mathrm{V}}) f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}}) \\ &\propto \theta_{\mathrm{V}}^{-n/2} e^{-\frac{1}{2\theta_{\mathrm{V}}}((n-1)s^2+n(\bar{x}-\theta_{\mathrm{M}})^2} \cdot \theta_{\mathrm{V}}^{-1/2} \theta_{\mathrm{V}}^{-\alpha_0-1} e^{-\frac{1}{2\theta_{\mathrm{V}}}(2\beta_0+\lambda_0(\theta_{\mathrm{M}}-\mu_0)^2)} \\ &= \theta_{\mathrm{V}}^{-1/2} \theta_{\mathrm{V}}^{-\alpha_0-n/2-1} e^{-\frac{1}{2\theta_{\mathrm{V}}}\left((n-1)s^2+n(\bar{x}-\theta_{\mathrm{M}})^2+2\beta_0+\lambda_0(\theta_{\mathrm{M}}-\mu_0)^2\right)}. \end{aligned}$$

Om vi utvecklar kvadrattermerna i exponenten kan vi skriva om dem som

$$\begin{aligned}
&n(\bar{x}-\theta_{\mathrm{M}})^2+\lambda_0(\theta_{\mathrm{M}}-\mu_0)^2 \\
&\quad= n\bar{x}^2-2n\bar{x}\theta_{\mathrm{M}}+n\theta_{\mathrm{M}}^2+\lambda_0\theta_{\mathrm{M}}^2-2\lambda_0\mu_0\theta_{\mathrm{M}}+\lambda_0\mu_0^2 \\
&\quad= (\lambda_0+n)\theta_{\mathrm{M}}^2+2(\lambda_0\mu_0+n\bar{x})\theta_{\mathrm{M}}+n\bar{x}^2+\lambda_0\mu_0^2 \\
&\quad= (\lambda_0+n)\left(\theta_{\mathrm{M}}^2+2\left(\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)\theta_{\mathrm{M}}+\left(\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2\right)-(\lambda_0+n)\left(\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2+n\bar{x}^2+\lambda_0\mu_0^2 \\
&\quad= (\lambda_0+n)\left(\theta_{\mathrm{M}}-\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2+\frac{-(\lambda_0\mu_0+n\bar{x})^2+(\lambda_0+n)n\bar{x}^2+(\lambda_0+n)\lambda_0\mu_0^2}{\lambda_0+n} \\
&\quad= (\lambda_0+n)\left(\theta_{\mathrm{M}}-\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2+\frac{-\lambda_0^2\mu_0^2-2n\bar{x}\lambda_0\mu_0-n^2\bar{x}^2+(\lambda_0+n)n\bar{x}^2+(\lambda_0+n)\lambda_0\mu_0^2}{\lambda_0+n} \\
&\quad= (\lambda_0+n)\left(\theta_{\mathrm{M}}-\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2+\frac{\lambda_0 n}{\lambda_0+n}\left(\bar{x}^2-2\bar{x}\mu_0+\mu_0^2\right) \\
&\quad= (\lambda_0+n)\left(\theta_{\mathrm{M}}-\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2+\frac{\lambda_0 n}{\lambda_0+n}\left(\bar{x}-\mu_0\right)^2.
\end{aligned}$$

(Notera att vi inte kan dra ifrån och lägga till konstanter på samma sätt som i andra fall då alla dessa termer multipliceras med $1/2\theta_{\mathrm{V}}$ i täthetsfunktionen och därför inte kan anses konstanta med avseende på $\theta_{\mathrm{M}},\theta_{\mathrm{V}}$.) Om vi stoppar in detta i uttrycket ovan för täthetsfunktionen får vi

$$f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}\mid\boldsymbol{X}}(\theta_{\mathrm{M}},\theta_{\mathrm{V}}\mid\boldsymbol{x})=\theta_{\mathrm{V}}^{-1/2}\theta_{\mathrm{V}}^{-\alpha_0-n/2-1}e^{-\frac{1}{2\theta_{\mathrm{V}}}\left(2\left(\beta_0+\frac{(n-1)s^2}{2}+\frac{\lambda_0 n}{\lambda_0+n}\frac{(\bar{x}-\mu_0)^2}{2}\right)+(\lambda_0+n)\left(\theta_{\mathrm{M}}-\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n}\right)^2\right)}.$$

Vi har alltså

$$\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}\mid\boldsymbol{X}=\boldsymbol{x}\sim\mathrm{NInvGamma}(\mu_n,\lambda_n,\alpha_n,\beta_n)$$

där

$$\begin{aligned}
\mu_n&=\frac{\lambda_0\mu_0+n\bar{x}}{\lambda_0+n},\\
\lambda_n&=\lambda_0+n,\\
\alpha_n&=\alpha_0+\frac{n}{2},\\
\beta_n&=\beta_0+\frac{n-1}{2}s^2+\frac{n\lambda_0}{\lambda_0+n}\frac{(\bar{x}-\mu_0)^2}{2}.
\end{aligned}$$

Som vi noterat tidigare är det ofta så att vissa parametrar är mer intressanta än andra. För normalfördelningen kanske vi har både okänt väntevärde Θ_{M} och varians Θ_{V}, men det är endast Θ_{M} som intresserar oss. I detta fall kan vi beräkna marginalfördelningen för Θ_{M}. Denna fås genom

$$f_{\Theta_{\mathrm{M}}\mid\boldsymbol{X}}(\theta_{\mathrm{M}}\mid\boldsymbol{x})\propto\int_0^{+\infty}\theta_{\mathrm{V}}^{-1/2}\theta_{\mathrm{V}}^{-\alpha_n-1}e^{-\frac{1}{2\theta_{\mathrm{V}}}\left(2\beta_n+\lambda_n(\theta_{\mathrm{M}}-\mu_n)^2\right)}d\theta_{\mathrm{V}}.$$

Om vi sätter $Q=2\beta_n+\lambda_n(\theta_{\mathrm{M}}-\mu_n)^2$ och gör variabelbytet $u=\frac{1}{2\theta_{\mathrm{V}}}Q$ får vi $d\theta_{\mathrm{V}}=\frac{1}{2u^2}Qdu$, vilket ger

$$\begin{aligned}
f_{\Theta_{\mathrm{M}}\mid\boldsymbol{X}}(\theta_{\mathrm{M}}\mid\boldsymbol{x})&\propto\int_0^{+\infty}\left(\frac{1}{2u}Q\right)^{-1/2}\left(\frac{1}{2u}Q\right)^{-\alpha_n-1}e^{-u}\frac{1}{2u^2}Qdu\\
&\propto Q^{-\alpha_n-1/2}\int_0^{+\infty}\int_0^{+\infty}u^{\alpha_n-1/2}e^{-u}du.
\end{aligned}$$

Integralen är konstant med avseende på θ_{M}, så vi får

$$\begin{aligned}
f_{\Theta_{\mathrm{M}}\mid\boldsymbol{X}}(\theta_{\mathrm{M}}\mid\boldsymbol{x})&\propto Q^{-\alpha_n-1/2}=\left(2\beta_n+\lambda_n(\theta_{\mathrm{M}}-\mu_n)^2\right)^{-\alpha_n-1/2}\\
&\propto\left(1+\frac{1}{2\alpha}\frac{(\theta_{\mathrm{M}}-\mu_n)^2}{\beta_n/\alpha_n\lambda_n}\right)^{-\frac{2\alpha_n+1}{2}}.
\end{aligned}$$

Formen på täthetsfunktionen sammanfaller med den generaliserade t-fördelningen med $2\alpha_n$ frihetsgrader skiftad med μ_n och skalad med $\sqrt{\beta_n/\alpha_n\lambda_n}$, dvs.

$$\Theta_{\mathrm{M}} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{t}\left(2\alpha_n, \mu_n, \frac{\beta_n}{\alpha_n\lambda_n}\right)$$

För Θ_{V} har vi

$$\Theta_{\mathrm{V}} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{InvGamma}(\alpha_n, \beta_n)$$

från definitionen av normal- och inversa gammafördelningen.

Från innan har vi sett att en icke-informativ apriorifördelning för Θ_{M} är $f_{\Theta_{\mathrm{M}}}(\theta_{\mathrm{M}}) \propto 1$ medan för Θ_{V} har vi $f_{\Theta_{\mathrm{V}}}(\theta_{\mathrm{V}}) \propto 1/\theta_{\mathrm{V}}$. Om vi antar att $\Theta_{\mathrm{M}} \perp\!\!\!\perp \Theta_{\mathrm{V}}$ får vi den icke-informativa apriorifördelningen

$$f_{\Theta_{\mathrm{M}},\Theta_{\mathrm{V}}}(\theta_{\mathrm{M}}, \theta_{\mathrm{V}}) \propto \frac{1}{\theta_{\mathrm{V}}}.$$

Aposteriorifördelningen blir då

$$\boldsymbol{\Theta} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{NInvGamma}(\bar{x}, n, (n-1)/2, (n-1)s^2/2).$$

Detta ger i sin tur de marginella aposteriorifördelningarna

$$\Theta_{\mathrm{M}} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{t}\left(n-1, \bar{x}, s^2/n\right)$$

och

$$\Theta_{\mathrm{V}} \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{InvGamma}((n-1)/2, (n-1)s^2/2).$$

Dessa resultat kan också generaliseras till vektorvärd data (dvs. där datafördelningen är en flerdimensionell normalfördelning). Vi ersätter då den inversa gammafördelningen med en annan fördelning över postitivt definita matriser som kallas den inversa Wishart fördelningen (se övning 7).

7.3.3 Exponentialfamiljer

Vi avslutar med ett resultat som sammanbinder exponentialfamiljer, tillräckliga statistikor, och naturligt konjugerade apriorifördelningar. Som vi sett innan så har vi en exponentialfamilj med avseende på $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^d$ om täthetsfunktionen för en stokastisk vektor $\boldsymbol{X} \in \mathbb{R}^m$ uppfyller

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = h(\boldsymbol{x})c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})^{\mathrm{T}} t_i(\boldsymbol{x})\right)$$

där $h : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$, $c : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ och $w_i : \mathbb{R}^d \to \mathbb{R}^{k_i}$ samt $t_i : \mathbb{R}^m \to \mathbb{R}^{k_i}$ för något positivt heltal k_i för alla $i = 1, 2, \ldots, k$. (Här är $\boldsymbol{X}$ en kontinuerlig stokastisk vektor, men följande resonemang håller även i det diskreta fallet.) Vi kan generalisera Sats 4.5 och visa att givet ett oberoende och likafördelat stickprov $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ så utgör

$$\left[\sum_{j=1}^{n} t_1(\boldsymbol{X}_j), \sum_{j=1}^{n} t_2(\boldsymbol{X}_j), \ldots, \sum_{j=1}^{n} t_k(\boldsymbol{X}_j)\right]^{\mathrm{T}}$$

en tillräcklig statistika för $\boldsymbol{\theta}$ (av dimension $(\sum_{i=1}^{k} k_i)$).

Vi kan också definiera naturliga parametrar i det flerdimensionella fallet. Detta ger parametriseringen $\boldsymbol{\eta} = [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_k]^{\mathrm{T}}$ där $\boldsymbol{\eta}_i \in \mathbb{R}^{k_i}$ och $\boldsymbol{\eta}$ uppfyller

$$\int_{\mathbb{R}^m} h(\boldsymbol{x}) \exp\left(\sum_{i=1}^{k} \boldsymbol{\eta}_i^{\mathrm{T}} t_i(\boldsymbol{x})\right) d\boldsymbol{x} < +\infty.$$

Vi har då faktoriseringen

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\eta}) = h(\boldsymbol{x})c(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^{k} \boldsymbol{\eta}_i^{\mathrm{T}} t_i(\boldsymbol{x})\right)$$

där $c(\boldsymbol{\eta})$ väljs så att täthetsfunktionen integrerar till ett.

Om vi nu antar att vi har ett stickprov $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ som är betingat oberoende givet $\boldsymbol{H} = \boldsymbol{\eta}$ och varje observation $\boldsymbol{X}_i$ har den betingade täthetsfunktionen ovan givet $\boldsymbol{H} = \boldsymbol{\eta}$, så söker vi den naturligt konjugerade familjen av apriorifördelningar (versalformen av η är H). Vi definierar då en exponentialfamilj i $\boldsymbol{H}$ genom

$$f_{\boldsymbol{H}}(\boldsymbol{\eta}) \propto c(\boldsymbol{\eta})^{\nu_0} \exp\left(\sum_{i=1}^{k} \boldsymbol{\eta}_i^{\mathrm{T}} \boldsymbol{\alpha}_{0,i}\right)$$

där $\nu_0 \in \mathbb{R}_{>0}$ och $\boldsymbol{\alpha}_{0,i} \in \mathbb{R}^{k_i}$ för $i = 1, 2, \ldots, k$ och tar denna som apriorifördelning. Detta ger då aposteriorifördelningen

$$\begin{aligned} f_{\boldsymbol{H} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n}(\boldsymbol{\eta} \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &\propto c(\boldsymbol{\eta})^{\nu_0 + n} \exp\left(\sum_{i=1}^{k} \boldsymbol{\eta}_i^{\mathrm{T}} \left(\boldsymbol{\alpha}_{0,i} + \sum_{j=1}^{n} t_i(\boldsymbol{x}_j)\right)\right) \\ &= c(\boldsymbol{\eta})^{\nu_n} \exp\left(\sum_{i=1}^{n} \boldsymbol{\eta}_i^{\mathrm{T}} \boldsymbol{\alpha}_{n,i}\right), \end{aligned}$$

där $\nu_n = \nu_0 + n$ och $\boldsymbol{\alpha}_{n,i} = \boldsymbol{\alpha}_{0,i} + \sum_{j=1}^{n} t_i(\boldsymbol{x}_j)$. Med andra ord bildar denna exponentialfamilj en naturligt konjugerad familj av apriorifördelningar för den naturliga parametern. Uppdateringsregeln för ν lägger till antalet observationer medan för de andra parametrarna $\boldsymbol{\alpha}_i$ lägger vi till de tillräckliga statistikorna $\sum_{j=1}^{n} t_i(\boldsymbol{x}_j)$.

En intressant observation är att aposteriorifördelningen här endast beror på stickprovsstorleken n samt de tillräckliga statistikorna för våra parametrar. Man kan visa att detta håller mer generellt (dvs. inte bara för exponentialfamiljer med den naturliga parametriseringen). När man söker aposteriorifördelningen för en ny datafördelning kan det därför vara användbart att beräkna aposteriorifördelningen med avseende på de tillräckliga statistikorna då detta kan avsevärt förenkla kalkylen.