

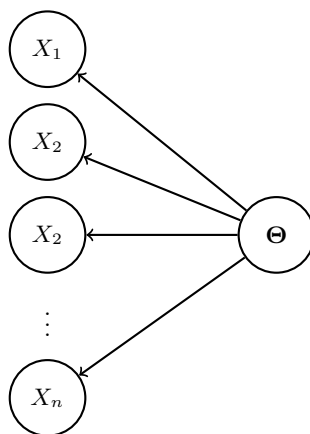


Föreläsning 14

Statistisk inläring och dataanalys (Kungliga Tekniska Högskolan)

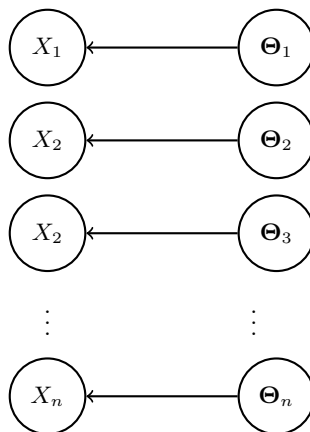
Föreläsning 14

Tidigare har vi skattat en eller ett par parametrar utifrån data med hjälp av metoder från bayesiansk inferens. Den vanligaste situationen är att vi har något stickprov X_1, X_2, \dots, X_n där varje observation är betingat oberoende med avseende på någon parametervektor Θ . Vidare har vi att de betingade datafördelningarna är identiska med avseende på denna underliggande parameter. I grafform får vi följande modell.

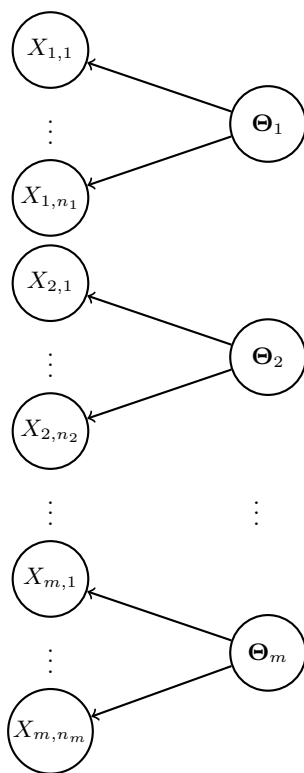


En viktig egenskap hos denna modell är att parametervektorn Θ är densamma för *alla* observationer. Med andra ord antar vi att observationerna är starkt kopplade till varandra (de kanske är repetitioner av samma experiment eller identiska produkter som undersöks på samma sätt). Vi kan variera modellen ovan genom att variera datafördelningarna mellan de olika observationerna, men den underliggande parametern Θ är alltid densamma.

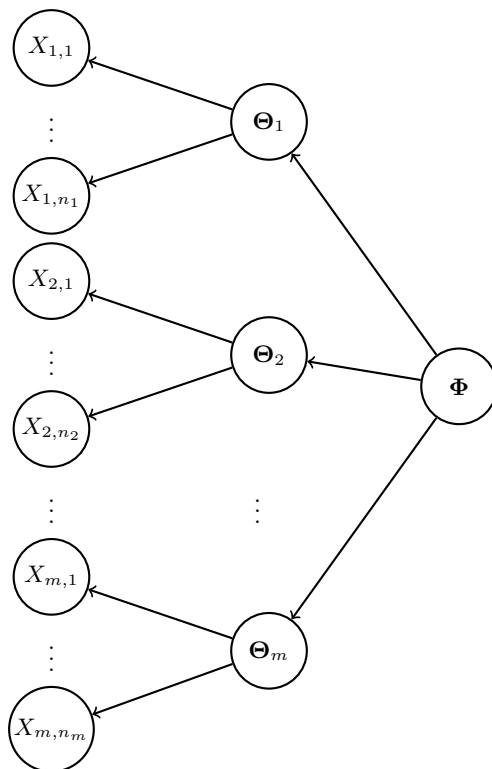
I många fall har vi dock inte alltid en lika stark koppling mellan observationerna. Vi har kanske en grupp experiment som utförs på olika sätt och beror på olika underliggande parametrar eller så har vi liknande produkter från olika fabriker som ska undersökas. En variant är då att ansätta en parameter Θ_j som hör till varje observation X_j , vilket ger följande modell.



Mer allmänt kanske vi har grupper av observationer $X_{j,1}, X_{j,2}, \dots, X_{j,n_j}$ för något $n_j \in \mathbb{Z}_{>0}$ som alla delar samma parametervektor Θ_j för $j = 1, 2, \dots, m$. Detta ger följande modell.



Vi kan nu skatta de olika parametrarna $\Theta_1, \Theta_2, \dots, \Theta_m$ var för sig (givet någon apriorifördelning för parametrarna). Nackdelen här är att vi saknar någon relation mellan de olika parametervektorerna. Även om de inte är identiska kanske det finns någon länk mellan dem. Ett vanligt antagande är att parametervektorerna i sig kan anses vara betingat oberoende givet en annan parametervektor Φ . Den gemensamma parametervektorn Φ kallas ofta för *hyperparameter* då de är parametrar för de ordinarie parametrarna $\Theta_1, \Theta_2, \dots, \Theta_m$. Vi får då följande modell.



Detta kallas för en *bayesiansk hierarkisk modell* (i två nivåer). Resten av föreläsningen kommer vi att ägna oss åt att skatta parametrar och hyperparametrar i denna typ av modeller.

14.1 Empiriska bayesmetoden

Ett sätt att angripa en bayesiansk hierarkisk modell är att konstruera punktskattningar för hyperparametrarna och sedan använda den resulterande apriorifördelningen för att genomföra traditionell bayesiansk inferens för varje parameter var för sig. Det finns en mängd olika sätt att göra detta med hjälp av ML-skattningar av hyperparametrarna eller momentmetoden. Dessa kallas för *empiriska bayesmetoder* då de använder observationerna (dvs. den empiriska datan) för att konstruera en apriorifördelning.

I denna kurs kommer vi att betrakta en specifik variant av den empiriska bayesmetoden. Metoden har följande steg.

1. Skatta var parameter Θ_j för sig med momentmetoden från datat $x_{j,1}, x_{j,2}, \dots, x_{j,n_j}$ för $j = 1, 2, \dots, m$. Beteckna skattningen med $\hat{\theta}_j$.
2. Dessa $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ betraktas nu som utfall av $\Theta_1, \Theta_2, \dots, \Theta_m$ och används för att skatta hyperparametrarna Φ med momentmetoden. Denna skattning betecknas $\hat{\phi}$.
3. Fixera $\Phi = \hat{\phi}$ i modellen och använd den resulterande fördelningen för Θ som apriorifördelning.
4. Givet nya datapunkter x_1, x_2, \dots, x_n , bestäm aposteriorifördelningen för Θ med denna apriorifördelning.

Eftersom metoden bygger på punktskattningar av hyperparametrarna och inte tar deras variation i beaktning (dvs. eftersom vi inte beräknar någon aposteriorifördelning med avseende på hyperparametrarna) så är inte denna modell bayesiansk i dess sanna bemärkelse. Den empiriska bayesmetoden är dock ett enkelt sätt att inkorporera tidigare observationer för att förbättra inferens i en situation med få mätdata. Vi illustrerar den med ett exempel.

Exempel 14.1. Ett antal råttor undersöks för utveckling av tumörer. Råttorna är av standardtypen "F344" kommer från en kontrollgrupp som inte har utsatts för någon intervention (t.ex. injektion av ett läkemedel). I vårt experiment observerar vi 14 råttor, varav 4 utvecklar tumörer. Vi skulle nu vilja bestämma den underliggande sannolikheten att utveckla tumörer för dessa råttor.

Denna observation X är binomialfördelad givet sannolikheten $\Theta = \theta$ att råttan utvecklar en tumör. Med andra ord har vi $X | \Theta = \theta \sim \text{Bin}(14, \theta)$. En naturlig apriorifördelning är betafördelningen, då denna familj är konjugerad med binomialfördelningen. Vi har då $\Theta \sim \text{Beta}(\alpha, \beta)$, men behöver parametrarna α och β för att göra inferens.

Ett tillvägagångssätt är att anta en icke-informativ apriorifördelning $\Theta \sim \text{Beta}(0, 0)$ (Haldanes apriorifördelning). Detta ger aposteriorifördelningen $\Theta | X = 4 \sim \text{Beta}(4, 10)$ och aposterioriväntevärdet $E[\Theta | X = 4] = 4/14 \approx 0.286$ med aposteriostandardavvikelsen $\text{Std}[\Theta | X = 4] \approx 0.117$. Med andra ord har vi ganska stor osäkerhet i vår skattning.

Låt oss nu anta att andra experiment har gjorts på samma typ av råttor (dvs. av typen F344) och följande resultat har uppnåtts

(0,20)	(0,20)	(0,20)	(0,20)	(0,20)	(0,20)	(0,20)	(0,19)	(0,19)	(0,19)
(0,19)	(0,18)	(0,18)	(0,17)	(1,20)	(1,20)	(1,20)	(1,20)	(1,19)	(1,19)
(1,18)	(1,18)	(2,25)	(2,24)	(2,23)	(2,20)	(2,20)	(2,20)	(2,20)	(2,20)
(2,20)	(1,10)	(5,49)	(2,19)	(5,46)	(3,27)	(2,17)	(7,49)	(7,47)	(3,20)
(3,20)	(2,13)	(9,48)	(10,50)	(4,20)	(4,20)	(4,20)	(4,20)	(4,20)	(4,20)
(4,20)	(10,48)	(4,19)	(4,19)	(4,19)	(5,22)	(11,46)	(12,49)	(5,20)	(5,20)
(6,23)	(5,19)	(6,22)	(6,20)	(6,20)	(6,20)	(16,52)	(15,47)	(15,46)	(9,24)

där beteckningen (x, n) motsvarar x råttor med tumörer av n totalt. Vi kan använda den empiriska bayesmetoden för att skatta hyperparametrarna α och β .

För varje experiment (x_j, n_j) har vi skattningen av $\theta_j = x_j/n_j$. Med andra ord har vi $\hat{\theta}_1 = 0/20 = 0$, $\hat{\theta}_2 = 0/20$, \dots , $\hat{\theta}_{70} = 9/24 = 0.375$. Om vi skattar medelvärdet och variansen får vi

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^m \hat{\theta}_i \approx 0.136$$

respektive

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^m (\hat{\theta}_j - \hat{\mu})^2 \approx 0.0106.$$

Om vi sätter formlerna för $E[\Theta_j]$ och $\text{Var}[\Theta_j]$ från betafördelningen lika med dessa får vi skattningen $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$.

Vi kan nu använda apriorifördelningen $\Theta \sim \text{Beta}(1.4, 8.6)$ för att ta fram aposteriorifördelningen för vår data $X = 4$ där $X \mid \Theta = \theta \sim \text{Bin}(14, \theta)$. Detta ger då $\Theta \mid X = 4 \sim \text{Beta}(1.4 + 5, 8.6 + 10) = \text{Beta}(5.4, 18.6)$ som ger aposterioriväntevärdet $5.4/(5.4 + 18.6) \approx 0.223$. Vi får alltså ett lägre aposterioriväntevärde än för den icke-informativa apriorifördelningen $\text{Beta}(0, 0)$. Detta följer från faktumet att antalet tumörer i vårt experiment är ovanligt högt jämfört med de andra experimenten, så sannolikheten justeras ned. Aposterioristandardavvikelsen $\text{Std}[\Theta \mid X = 4] \approx 0.083$ är också mindre, vilket representerar en större tillförlitlighet i skattningen.

Ett tredje alternativ är att anta att alla experiment har samma Θ , dvs. $X_j \mid \Theta = \theta \sim \text{Bin}(n_j, \theta)$ för alla $j = 1, 2, \dots, 70$. I detta fall kan vi anta att Θ har en icke-informativ apriorifördelning $\Theta \sim \text{Beta}(0, 0)$. Detta skulle ge ett aposterioriväntevärde $E[\Theta \mid X_1 = x_1, \dots, X_{70} = x_{70}] \approx 0.150$ och aposterioristandardavvikelse $\text{Std}[\Theta \mid X_1 = x_1, \dots, X_{70} = x_{70}] \approx 0.00865$. Väntevärdet är alltså mycket mindre och likaså standardavvikelsen. Här tillåter vi alltså ingen variation mellan de olika kontrollgrupperna och får då en mindre skattning som i sin tur också är ganska säker (dvs. har låg standardavvikelse). Olika antaganden ger alltså olika resultat i detta fall. Vilken skattning som är bäst beror hur väl modellen representerar verkligheten (detta kan, till exempel utvärderas med aposterioriprediktiv validering eller informationskriterier).

14.2 Bayesiansk inferens i en binomialhierarki

Den empiriska bayesmetoden är som sagt ett sätt genomföra inferens i en hierarkisk modell. Problemet är att vi ansätter fixa värden för hyperparametrarna, vilket gör att metoden inte ger hela bilden från ett bayesianskt perspektiv. Punktskattningarna $\hat{\phi}$ för Φ måste då ersättas med aposteriorifördelningar som vi sedan kan generera utfall från och beräkna olika värden.

Om alla apriorifördelningar är kända kan vi göra detta med de verktyg vi har tagit fram hittills. Det viktiga är att kunna skilja på de betingade aposteriorifördelningarna (där vi håller en parameter eller hyperparameter fix) och de marginaliserade aposteriorifördelningarna (där vi integrerar bort en parameter eller hyperparameter). Om vi sätter ihop dessa på rätt sätt kan vi sedan få ut de skattningar vi är intresserade av. Till skillnad från den empiriska bayesmetoden innefattar denna mer kompletta metod en hel del mer beräkning, men resultatet är en mer flexibel metod som har större chans att ge oss en bättre insikt i datat.

Låt oss anta att vi, som innan, har observationer $X_{j,1}, X_{j,2}, \dots, X_{j,n_j}$ betingat oberoende givet Θ_j där $n_j \in \mathbb{Z}_{>0}$ med samma betingade fördelningar för $j = 1, 2, \dots, m$. Vidare har vi att $\Theta_1, \Theta_2, \dots, \Theta_m$ är betingat oberoende givet någon hyperparameter Φ med samma betingade fördelningar. (Vi har också att $X_{j,i} \perp X_{j',i'} \mid \Phi$).

Anta nu att vi har, för varje $j = 1, 2, \dots, m$, observationen $[X_{j,1}, \dots, X_{j,n_j}]^T = [x_{j,1}, \dots, x_{j,n_j}]^T$, vilket vi betecknar med $\mathbf{X}_j = \mathbf{x}_j$. Om vi samlar alla observationer i en enda stor vektor kan vi skriva detta som $\mathbf{X} = \mathbf{x}$. Som vanligt kan vi bilda täthetsfunktionen för aposteriorifördelningen

$$f_{\Theta, \Phi | \mathbf{X}}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{x}) \propto f_{\mathbf{X} | \Theta, \Phi}(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) f_{\Theta, \Phi}(\boldsymbol{\Theta}, \boldsymbol{\Phi}),$$

där vi har använt $\boldsymbol{\Theta}$ för att beteckna vektorn $[\Theta_1, \dots, \Theta_m]^T$. Eftersom vi har den betingade fördelningen för $\boldsymbol{\Theta}$ givet $\boldsymbol{\Phi}$ skriver vi om detta som

$$f_{\Theta, \Phi | \mathbf{X}}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{x}) \propto f_{\mathbf{X} | \Theta, \Phi}(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) f_{\boldsymbol{\Theta} | \boldsymbol{\Phi}}(\boldsymbol{\Theta} \mid \boldsymbol{\Phi}) f_{\boldsymbol{\Phi}}(\boldsymbol{\phi}).$$

Formellt sett räcker det med resultatet ovan. Om vi sätter in de observerade datapunkterna \mathbf{x} och en passande apriorifördelning $f_{\boldsymbol{\Phi}}(\boldsymbol{\phi})$ för hyperparametrarna kan vi sedan använda MCMC-metoder för att generera utfall från $(\boldsymbol{\Theta}, \boldsymbol{\Phi})$ och beräkna olika väntevärden. Problemet är att aposteriorifördelningen ovan, som kallas *den simultana aposteriorifördelningen* eftersom den bestämmer fördelningen för $\boldsymbol{\Theta}$ och $\boldsymbol{\Phi}$ simultant, kan vara svår att simulera ifrån då antalet parametrar ofta är stort. Ett annat problem är att det kan vara svårt att få insikt i hur de olika parametrarna påverkar varandra utifrån detta uttryck.

Nästa steg är därför att ta fram den betingade aposteriorifördelningen för $\boldsymbol{\Theta}$ givet \mathbf{X} och $\boldsymbol{\Phi}$, dvs. aposteriorifördelningen för $\boldsymbol{\Theta}$ om vi antar att hyperparametern $\boldsymbol{\Phi}$ är fix. Detta motsvarar sista steget i den empiriska bayesmetoden och är oftast ganska enkel om vi väljer $\boldsymbol{\Theta}$ s fördelning från en familj av konjugerade apriorifördelningar med

avseende på datafördelningarna. Vi kan då använda standardreglerna för bayesiansk uppdatering. Utifrån detta får vi ett uttryck för $f_{\Theta|\Phi, \mathbf{X}}(\theta | \phi, \mathbf{x})$.

Det sista steget är att ta fram den marginaliserade aposteriorifördelningen för Φ givet \mathbf{X} . Ett sätt att ta fram detta är att integrera den simultana aposteriorifördelningen $f_{\Theta, \Phi|\mathbf{X}}(\theta, \phi | \mathbf{x})$ med avseende på θ . Detta kräver ofta ganska mycket jobb, så vi använder oss istället av ett litet trick. Enligt kedjeregeln har vi

$$f_{\Theta, \Phi|\mathbf{X}}(\theta, \phi | \mathbf{x}) = f_{\Theta|\Phi, \mathbf{X}}(\theta | \phi, \mathbf{x}) f_{\Phi|\mathbf{X}}(\phi | \mathbf{x}),$$

vilket vi kan skriva om som

$$f_{\Phi|\mathbf{X}}(\phi | \mathbf{x}) = \frac{f_{\Theta, \Phi|\mathbf{X}}(\theta, \phi | \mathbf{x})}{f_{\Theta|\Phi, \mathbf{X}}(\theta | \phi, \mathbf{x})}.$$

Med andra ord kan vi helt enkelt stoppa in den simultana aposteriorifördelningen för Θ, Φ givet \mathbf{X} och den betingade aposteriorifördelningen för Θ givet Φ, \mathbf{X} i formeln ovan och få ut den marginaliserade aposteriorifördelningen för Φ givet \mathbf{X} . Observera att man måste vara försiktig med täthetsfunktionen i täljaren. I vanliga fall kan vi komma undan genom att ta bort olika konstanter i täthetsfunktionen som inte beror på Θ , men i detta fall kan dessa vara viktiga (om de beror på Φ) när vi använder dem i formeln ovan. Därför är det säkrast att arbeta med normaliserade täthetsfunktioner för den betingade aposteriorifördelningen (vilket ofta inte är något problem när vi arbetar med konjugerade apriorifördelningar då dessa ger oss de nödvändiga konstanterna).

Med denna marginaliserade aposteriorifördelning för hyperparametrarna kan vi sedan utföra inferens genom sampling från de olika fördelningarna. Först genererar vi utfall från den aposteriorifördelningen för hyperparametrarna Φ , sedan använder vi dessa för att generera utfall för parametrarna $\Theta_1, \Theta_2, \dots, \Theta_m$ och beräknar de väntevärden som är av intresse.

Vi illustrerar detta med tillämpning på en binomialhierarki där $X_j | \Theta_j = \theta_j \sim \text{Bin}(n_j, \theta_j)$ och $\Theta_j | A = \alpha, B = \beta \sim \text{Beta}(\alpha, \beta)$ för $j = 1, 2, \dots, m$. Här har vi att $X_j \perp\!\!\!\perp X_{j'} | A, B$, $\Theta_j \perp\!\!\!\perp \Theta_{j'} | A, B$ samt $X_j \perp\!\!\!\perp \Theta_{j'} | A, B$ för $j \neq j'$. Vi väljer inte någon apriorifördelning för A, B än utan låter denna vara obestämd tills vidare.

Detta ger nu den simultana aposteriorifördelningen

$$f_{\Theta, A, B|\mathbf{X}}(\theta, \alpha, \beta | \mathbf{x}) \propto f_{A, B}(\alpha, \beta) f_{\Theta|A, B}(\theta | \alpha, \beta) f_{\mathbf{X}|\Theta, A, B}(\mathbf{x} | \theta, \alpha, \beta)$$

$$f_{A, B}(\alpha, \beta) \prod_{j=1}^m \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1 - \theta_j)^{\beta-1} \prod_{j=1}^m \binom{n_j}{x_j} \theta_j^{x_j} (1 - \theta_j)^{n_j - x_j}.$$

Eftersom de j olika beta-binomialhierarkierna $X_j | \Theta_j$ är betingat oberoende givet A, B har vi att $\Theta_j | X_j = x_j, A = \alpha, B = \beta \sim \text{Beta}(\alpha + x_j, \beta + n_j - x_j)$, dvs.

$$f_{\Theta_j|A, B, X_j}(\theta_j | \alpha, \beta, x_j) = \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + x_j)\Gamma(\beta + n_j - x_j)} \theta_j^{\alpha + x_j - 1} (1 - \theta_j)^{\beta + n_j - x_j - 1}.$$

Med alla $j = 1, 2, \dots, m$ tillsammans får vi då den betingade aposteriorifördelningen

$$f_{\Theta|A, B, \mathbf{X}}(\theta | \alpha, \beta, \mathbf{x}) = \prod_{j=1}^m \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + x_j)\Gamma(\beta + n_j - x_j)} \theta_j^{\alpha + x_j - 1} (1 - \theta_j)^{\beta + n_j - x_j - 1}.$$

Genom att använda vårt trick ovan får vi nu ut

$$f_{A, B|\mathbf{X}}(\alpha, \beta | \mathbf{x}) \propto f_{A, B}(\alpha, \beta) \prod_{j=1}^m \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + x_j)\Gamma(\beta + n_j - x_j)}{\Gamma(\alpha + \beta + n_j)}.$$

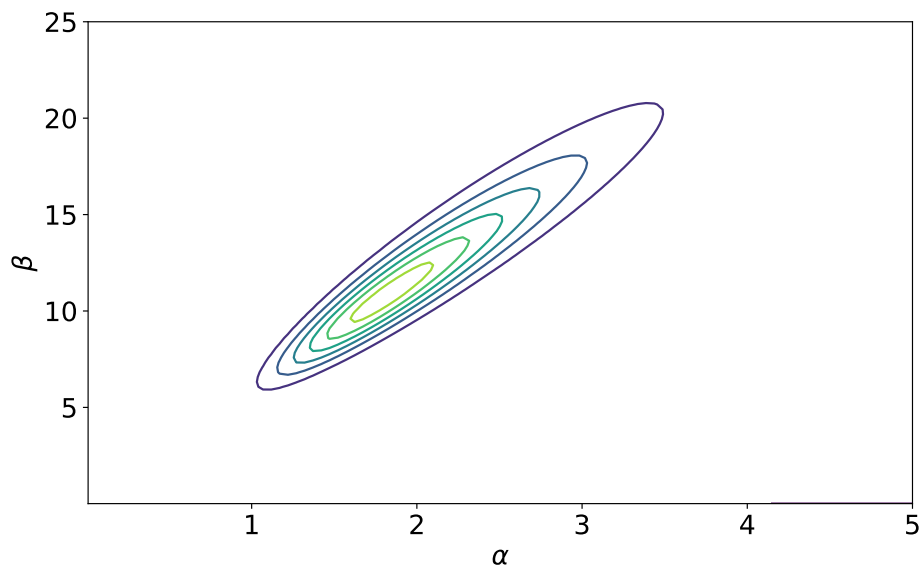
Denna formel kan vi inte förenkla ytterligare men går att beräkna numeriskt för godtyckliga α, β .

Valet av apriorifördelning för A, B återstår. Ett alternativ är att använda en egentlig apriorifördelning, vilket garanterar att aposteriorifördelningen också är egentlig. Att välja en bra icke-informativ apriorifördelning är svårt, men det går att visa att

$$f_{A, B}(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$$

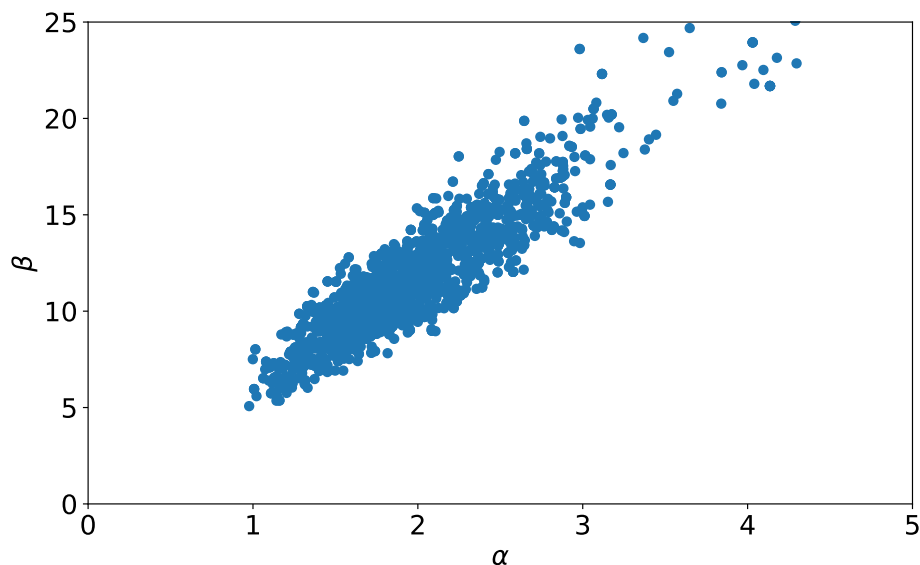
ger en egentlig aposteriorifördelning om det håller att $0 < x_j < n_j$ för något j . Med denna kan vi simulera utfall från A, B och sedan använda dessa för att simulera utfall av $\Theta_1, \Theta_2, \dots, \Theta_m$.

Exempel 14.2. Vi återgår till exemplet med råttorna. Om vi plottar aposteriorifördelningen för hyperparametrarna A och B får vi följande figur:



Fördelningen verkar alltså koncentrera sig kring $\alpha = 2$ och $\beta = 10$, vilket stämmer bra med punktskattningarna från den empiriska bayesmetoden (där vi fick $\hat{\alpha} = 1.4$ och $\hat{\beta} = 8.6$).

Vi kan simulera denna fördelning med hjälp av metropolisalgoritmen som vi har sett i tidigare. De 2000 utfall vi får från detta ser ut att följa fördelningen ganska bra:

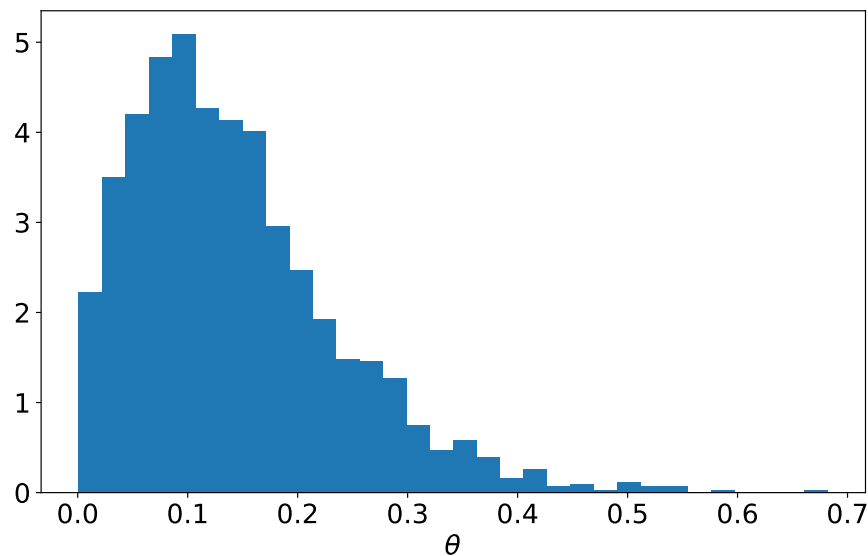


Om vi tar medelvärdet över alla dessa utfall kan vi approximera aposterioriväntevärdena för hyperparametrarna A och B , vilket ger

$$E[A \mid \mathbf{X} = \mathbf{x}] \approx 2.00 \quad \text{och} \quad E[B \mid \mathbf{X} = \mathbf{x}] \approx 11.9.$$

Detta är återigen nära punktskattningen vi fick från den empiriska bayesmetoden.

Vi kan också simulera fördelningen för Θ där $\Theta \mid A = \alpha, B = \beta \sim \text{Beta}(\alpha, \beta)$ där A och B följer fördelningen ovan. Detta görs enklast genom att simulera utfall från A, B och sedan använda dessa för att simulera utfall från Θ enligt betafördelningen. Vi får då histogrammet



Vi kan också använda dessa för att beräkna aposterioriväntevärdet och aposteriorivariansen för Θ . Ett bättre resultat kan vi få genom att utnyttja lagen om total förväntan och lagen om total sannolikhet. Detta ger

$$E[\Theta | \mathbf{X}] = E[E[\Theta | A, B, \mathbf{X}] | \mathbf{X}] = E\left[\frac{A}{A+B} \middle| \mathbf{X}\right] \approx 0.145$$

och

$$\begin{aligned} \text{Var}[\Theta | \mathbf{X}] &= E[\text{Var}[\Theta | A, B, \mathbf{X}] | \mathbf{X}] + \text{Var}[E[\Theta | A, B, \mathbf{X}] | \mathbf{X}] \\ &= E\left[\frac{AB}{(A+B)^2(A+B+1)} \middle| \mathbf{X}\right] + \text{Var}\left[\frac{A}{A+B} \middle| \mathbf{X}\right] \approx 0.00909 \end{aligned}$$

Båda är återigen ganska nära värdena som vi fick i den empiriska bayesmetoden, med en aningen mindre varians.

Med de simulerade utfallen från A, B kan vi använda den betingade aposteriorifördelningen för Θ_{71} för att ge en skattning av denna parameter som var aktuell i vårt experiment. Vi har då $\Theta_{71} | A = \alpha, B = \beta, \mathbf{X} = \mathbf{x} \sim \text{Beta}(\alpha + x_{71}, \beta + n_{71} - x_{71})$. Återigen använder vi lagarna om total förväntan och varians för att ta fram väntevärdet

$$E[\Theta_{71} | \mathbf{X}] = E[E[\Theta | A + X_{71}, B + n_{71} - X_{71}, \mathbf{X}] | \mathbf{X}] = E\left[\frac{A + X_{71}}{A + B + n_{71}} \middle| \mathbf{X}\right] \approx 0.217$$

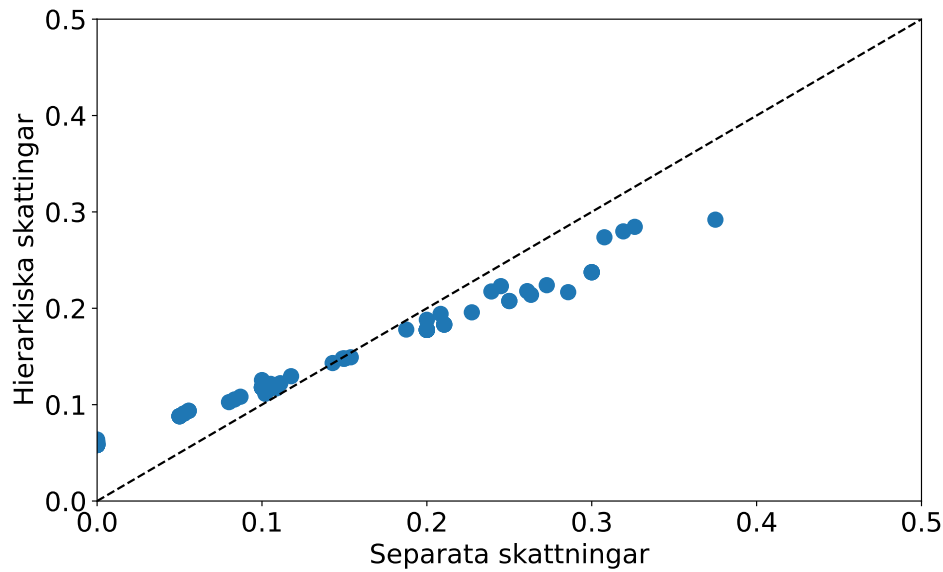
och variansen

$$\begin{aligned} \text{Var}[\Theta_{71} | \mathbf{X}] &= E[\text{Var}[\Theta_{71} | A, B, \mathbf{X}] | \mathbf{X}] + \text{Var}[E[\Theta_{71} | A, B, \mathbf{X}] | \mathbf{X}] \\ &= E\left[\frac{(A + X_{71})(B + n_{71} - X_{71})}{(A + B + n_{71})^2(A + B + n_{71} + 1)} \middle| \mathbf{X}\right] + \text{Var}\left[\frac{A + X_{71}}{A + B + n_{71}} \middle| \mathbf{X}\right] \approx 0.00616, \end{aligned}$$

vilket ger en aposterioristandardavvikelse på ungefär $\sqrt{0.00616} \approx 0.0785$. Väntevärdet ligger strax under skattningen från den empiriska bayesmetoden och standardavvikelsen är aningen mindre. Den mer kompletta bayesianska analysen har alltså reducerat väntevärdet mer mot de andra värden (som hade mycket färre tumörer jämfört med vårt experiment) och har gjort detta med lite större säkerhet. Detta kan bero på att aposteriorifördelningen för Θ (obetingat med avseende på datat) generellt är mer koncentrerad, vilket leder till större koncentration för Θ_{71} .

Återigen visar det sig att den empiriska bayesmetoden ger hyfsat bra resultat, då denna ligger ganska nära skattningarna som vi får med den kompletta modellen. Detta beror på att vi i denna situation har ganska ordentligt med data, så punktskattningen är av hög kvalitet och utgör en bra approximering av aposteriorifördelningen hos hyperparametern. Utan att göra denna närmare analys kan det dock vara svårt att komma fram till denna slutsats, så om man söker hög tillförlitlighet är det alltid en bra idé att försöka modellera hela hierarkin från ett bayesianskt perspektiv.

Avslutningsvis jämför vi skattningarna med skattningarna vi fick från de separata skattningarna x_j/n_j . Vi kan repetera simuleringarna ovan för alla parametrar $\Theta_1, \Theta_2, \dots, \Theta_{71}$ och ta fram deras aposterioriväntevärden. Om vi plottar dessa tillsammans med de separata skattningarna får vi



För experiment som hade högt antal tumörer ger den hierarkiska modellen en mindre skattning, medan för experiment med mycket få tumörer ger modellen en aningen större skattning. Vi krymper alltså variationen mot mitten av fördelningen för Θ , vilket är effekten av att modellen "binder ihop" parametrarna med hjälp av hyperparametrarna A och B .