Föreläsning 08

Statistisk inlärning och dataanalys (Kungliga Tekniska Högskolan)

Studocu är inte sponsrat och får inte stöd från något college eller universitet
Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

Föreläsning 8

8.1 Bayesiansk beslutsteori

Vi har tidigare sett hur beslutsteori kan tillämpas inom det frekventistiska ramverket för att utvärdera olika punktskattningar. Där hade vi en förlustfunktion $L(\boldsymbol{\theta}, \boldsymbol{a})$ som anger hur önskvärd en viss handling $\boldsymbol{a} \in \mathcal{A}$ är för ett visst värde $\boldsymbol{\theta} \in \Omega$ på parametern. I det fallet hade vi $\mathcal{A} = \Omega$ och $\boldsymbol{a} = \hat{\theta} = W(\boldsymbol{x})$ avsåg en skattning av $\boldsymbol{\theta}$.

Vi kan göra detsamma i det bayesianska ramverket. Givet datat observationen $X = x$ kan vi ta fram aposteriorifördelningen $\boldsymbol{\Theta} \mid X = x$ och använda denna för att beräkna en punktskattning, som till exempel aposteriorivärdet $W(X) = \mathrm{E}[\boldsymbol{\Theta} \mid X]$. Som innan kan vi då beräkna riskfunktionen, dvs. väntevärdet av förlustfunktionen över alla möjliga observationer, men för samma värde på parametern $\boldsymbol{\Theta}$. Vi har alltså ett betingat väntevärde:

$$R(\boldsymbol{\theta}, W) = \mathrm{E}[L(\boldsymbol{\Theta}, W(\boldsymbol{X})) \mid \boldsymbol{\Theta} = \boldsymbol{\theta}].$$

Man kan då variera $\boldsymbol{\theta}$ för att se hur skattningen beter sig och jämföra med andra skattningar.

Exempel 8.1. Antag att $X \mid \Theta = \theta \sim \mathrm{Bin}(n, \theta)$. Då har vi ML-skattningen

$$\hat{\Theta}_{\mathrm{ML}} = W_{\mathrm{ML}}(X) = \frac{X}{n}.$$

Givet en apriorifördelning $\mathrm{Beta}(\alpha, \beta)$ för Θ har vi också ett aposteriorivärde

$$\hat{\Theta}_{\mathrm{Bayes}} = W_{\mathrm{Bayes}}(X) = \frac{\alpha + X}{\alpha + \beta + n}.$$

Vi kan visa (prova!) att för kvadratfelet $L(\theta, a) = (\theta - a)^2$ har dessa skattningar riskfunktionerna

$$R(\theta, W_{\mathrm{ML}}) = \frac{\theta(1-\theta)}{n}$$

respektive

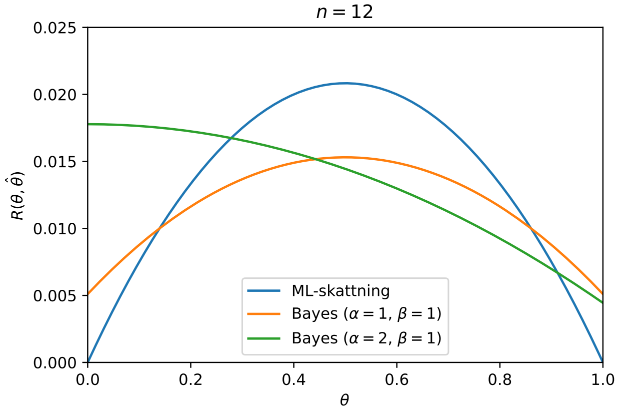
$$R(\theta, W_{\mathrm{Bayes}}) = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left(\frac{n\theta + \alpha}{\alpha+\beta+n} - \theta\right)^2.$$

Riskfunktionen för ML-skattningen är stor för θ runt $1/2$ och avtar för θ nära noll eller ett (för dessa värden har X mindre varians, vilket leder till ett mindre medelkvadratfel).

För aposteriorivärdet beror riskfunktionen såklart på valet av α och β. Om $\alpha = \beta = 0$ (dvs. en icke-informativ apriorifördelning) återfår vi ML-skattningen och dess riskfunktion. Låter vi α vara stort och β litet får vi ett stort medelkvadratfel för θ nära noll och ett litet för θ nära ett (apriorifördelningen är då koncentrerad runt ett, vilket ger bättre skattningar där).

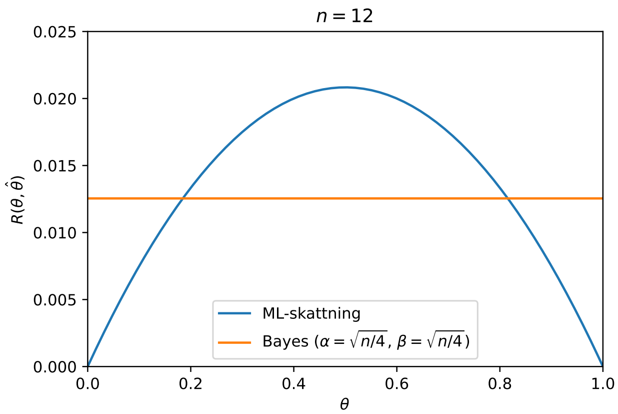
This document is available free of charge on studocu

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

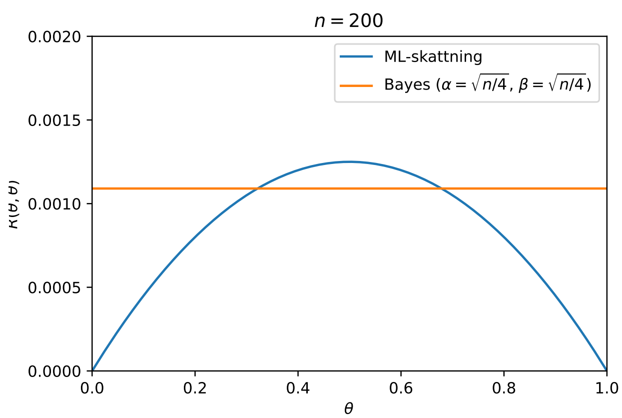


Ett speciellt val av aprioriparametrar α och β är $\alpha = \beta = \sqrt{n/4}$, vilket ger en konstant riskfunktion (verifiera detta!). Vi har då

$$R(\theta, W_{\text{Bayes}}) = \frac{n}{4(n + \sqrt{n})^2}.$$



När n är litet ger detta ett bättre medelkvadratfel för aposteriorivärdet när θ är runt $1/2$ jämfört med ML-skattningen. Vi ser dock att när $n \to +\infty$ så går medelkvadratfelet mot noll som $1/\sqrt{n}$ medan felet för ML-skattningen går mot noll som $1/n$, dvs. snabbare. För stora värden på n är då ML-skattningen att föredra, oavsett värdet på θ.



8.1.1 Bayesrisk och bayesskattning

Hittills har vi endast tillämpat verktygen från frekventistisk beslutsteori på våra bayesianska punktskattningar. Dessa ger riskfunktionen som beror på parametervärdet $\boldsymbol{\theta}$ och vi kan jämföra dessa. En viktig skillnad i det

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

bayesianska ramverket är att vi antar att $\boldsymbol{\theta}$ är ett utfall av en stokastisk vektor $\boldsymbol{\Theta}$ som har en viss fördelning – apriorifördelningen. Vi kan då sammanfatta risken hos en viss skattning genom att ta väntevärdet med avsende på denna fördelning. Vi får då

$$\mathrm{E}[R(\boldsymbol{\Theta}, W)] = \int_\Omega R(\boldsymbol{\theta}, W) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Detta värde kallas *bayesrisken* och motsvarar den genomsnittliga risken som vi får om vi repeterar skattningen för olika värden på $\boldsymbol{\theta}$ enligt apriorifördelningen av $\boldsymbol{\Theta}$.

Eftersom vi nu har sammanfattat utvärderingen av punkskattningen i ett enda värde kan detta användas för att rangordna olika punktskattningar (detta går inte med riskfunktionen då denna kan vara både större eller mindre för olika värden av $\boldsymbol{\theta}$). Om vi tror på vår apriorifördelning och vår förlustfunktion ges bayesrisken en bra fingervisning om vilken punktskattning som strikt bättre än annan. Bland annat kan vi söka efter den punktskattning som ger *lägst* bayesrisk. Denna skattning kallas för *bayesskattningen* med avseende på förlustfunktionen $L(\theta, a)$ och apriorifördelningen $f_\Theta(\theta)$.

Att beräkna bayesskattningen kan verka komplicerat. Först måste vi beräkna väntevärdet för förlustfunktionen över olika utfall $\boldsymbol{x}$ av $\boldsymbol{X}$ givet en fix parameter $\boldsymbol{\Theta} = \boldsymbol{\theta}$ med hjälp av datafördelningen. Sedan tar vi denna och bildar väntevärdet över olika utfall $\boldsymbol{\theta}$ av $\boldsymbol{\Theta}$ med hjälp av apriorifördelningen. Till slut ska vi optimera detta värde med avseende på själva skattningen, dvs. en funktion $W(\boldsymbol{x})$ från $\mathbb{R}^m$ till $\mathbb{R}^d$.

Det finns en (mycket) enklare väg att gå. Vi kan utnyttja faktumet att de två väntevärdena går att beräkna i en annan ordning. Mer specifikt har vi

$$\begin{aligned}
\mathrm{E}[R(\boldsymbol{\Theta}, W)] &= \int_\Omega R(\boldsymbol{\theta}, W) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_\Omega \left(\int_{\mathbb{R}^m} L(\boldsymbol{\theta}, W(\boldsymbol{x})) f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) d\boldsymbol{x} \right) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_{\mathbb{R}^m} \left(\int_\Omega L(\boldsymbol{\theta}, W(\boldsymbol{x})) f_{\boldsymbol{\Theta}|\boldsymbol{X}}(\boldsymbol{\theta} \mid \boldsymbol{x}) d\boldsymbol{\theta} \right) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x},
\end{aligned}$$

där vi har bytt integrationsordning och använt att $f_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x} \mid \boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = f_{\boldsymbol{\Theta}|\boldsymbol{X}}(\boldsymbol{\theta} \mid \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})$. Eftersom $f_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$ för alla $\boldsymbol{x}$ minimeras hela uttrycket om vi lyckas minimera den inre integralen för varje $\boldsymbol{x} \in \mathbb{R}^m$.

Denna inre integral kallas för *aposterioririsken* och är lika med

$$\mathrm{E}[L(\boldsymbol{\Theta}, W(\boldsymbol{X})) \mid X = x].$$

Detta värde motsvarar felet som begås i genomsnitt för skattningen av $\boldsymbol{\Theta}$ givet en viss observation $X = x$. Vi har alltså reducerat problemet med att hitta bayesskattningen till att minimera aposterioririsken för varje $x \in \mathbb{R}^m$. Eftersom $W(\boldsymbol{x})$ kan anta ett godtyckligt värde för varje $\boldsymbol{x}$ betyder detta att vi helt enkelt tar

$$W(\boldsymbol{x}) = \operatorname{argmin}_{\boldsymbol{a} \in \mathbb{R}^d} \mathrm{E}[L(\boldsymbol{\Theta}, \boldsymbol{a}) \mid X = x].$$

Detta minimeringsproblem är oftast betydligt enklare än originalproblemet och löses vanligast genom standardverktyg från differentialkalkylen.

Exempel 8.2. Antag att vi har kvadratfelet $L(\theta, a) = (\theta - a)^2$. Aposterioririsken är då

$$\begin{aligned}
\mathrm{E}[L(\theta, a) \mid X = x] &= \int_\Omega L(\theta, a) f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_\Omega (\theta - a)^2 f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_\Omega (\theta^2 - 2a\theta + a^2) f_{\Theta|X}(\theta \mid x) d\theta \\
&= \mathrm{E}[\Theta^2 \mid X = x] - 2a\, \mathrm{E}[\Theta \mid X = x] + a^2.
\end{aligned}$$

Denna funktion är minimierad i $a = \mathrm{E}[\Theta \mid X = x]$, så vår bayeskattning är

$$\mathrm{E}[\Theta \mid X],$$

dvs. aposteriorriväntevärdet.

Det visar sig alltså att om vi endast är intresserade av att minimera kvadratfelet till det sanna värdet θ så ges den optimala skattningen av aposterioriväntevärdet.

This document is available free of charge on studocu

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

Exempel 8.3. Om vi istället tar absolutfelet $L(\theta, a) = |\theta - a|$ och antar att $\Omega = \mathbb{R}$ får vi aposterioririsken

$$\begin{aligned}
\mathrm{E}[L(\theta,a) \mid X = x] &= \int_\Omega L(\theta,a) f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_\Omega |\theta - a| f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_{-\infty}^{a} |\theta - a| f_{\Theta|X}(\theta \mid x) d\theta + \int_{a}^{+\infty} |\theta - a| f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_{-\infty}^{a} (a - \theta) f_{\Theta|X}(\theta \mid x) d\theta + \int_{a}^{+\infty} (\theta - a) f_{\Theta|X}(\theta \mid x) d\theta \\
&= a \int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta - \int_{-\infty}^{a} \theta f_{\Theta|X}(\theta \mid x) d\theta + \int_{a}^{+\infty} \theta f_{\Theta|X}(\theta \mid x) d\theta - a \int_{a}^{+\infty} f_{\Theta|X}(\theta \mid x) d\theta.
\end{aligned}$$

Vi vill nu minimera med avseende på a och differentierar därför uttrycket med hjälp av integralkalkylens fundamentalsats ($\frac{d}{da}\int_b^a f(x)dx = f(a)$ och $\frac{d}{da}\int_a^b f(x)dx = -f(a)$). Detta ger

$$\begin{aligned}
\frac{d}{da} \mathrm{E}[L(\theta,a) \mid X = x] &= \int_{-\infty}^{a} f_{\Theta|X}(a \mid x) d\theta + a f_{\Theta|X}(a \mid x) - a f_{\Theta|X}(a \mid x) - a f_{\Theta|X}(a \mid x) - \int_{a}^{+\infty} f_{\Theta|X}(\theta \mid x) d\theta + a f_{\Theta|X}(a \mid x) \\
&= \int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta - \int_{a}^{+\infty} f_{\Theta|X}(\theta \mid x) d\theta \\
&= \int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta - \left(1 - \int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta\right) \\
&= 2 \int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta - 1.
\end{aligned}$$

Om vi sätter detta till noll får vi

$$\int_{-\infty}^{a} f_{\Theta|X}(\theta \mid x) d\theta = \frac{1}{2}.$$

Vi söker alltså a så att aposteriorifördelningens fördelningsfunktion uppfyller $F_{\Theta|X}(a \mid x) = 1/2$, dvs. aposteriorimedianen givet $X = x$.

Om vi differentierar en gång till får vi

$$\frac{d^2}{da^2} \mathrm{E}[L(\theta,a) \mid X = x] = 2 f_{\Theta|X}(a \mid x).$$

Denna andraderivata är positiv om vi har en unik median. Annars har vi ett intervall av parametervärden som alla är medianer till aposteriorifördelningen. Oavsett har vi hittat ett värde på a som minimerar aposterioririsken.

Exempel 8.4. Absolutfelet är ett specialfall av förlustfunktionen

$$L(\theta, a) = \begin{cases} c_0(\theta - a), & \theta - a \geq 0, \\ c_1(a - \theta), & \theta - a < 0, \end{cases}$$

för $c_0, c_1 \in \mathbb{R}_{>0}$ (absolutfelet återfås om vi tar $c_0 = c_1 = 1$. Denna förlustfunktion kallas för kvantilfel eller pingpongfel. Man kan visa att bayesskattningen med avseende på denna förlustfunktion ges av $c_1/(c_0 + c_1)$-kvantilen av aposteriorifördelningen, dvs. $W(x) = a$ för det a som uppfyller $F_{\Theta|X}(a \mid x) = 1 - c_1/(c_0 + c_1) = c_0/(c_0 + c_1)$.

Observera att i fallen ovan kan vi ta fram en sluten form för bayesskattningen, men detta är inte alltid möjligt för godtyckliga förlustfunktioner. I detta fall kan aposterioririsken behöva minimeras numeriskt. Om det finns en sluten form för aposterioririsken kan standardmetoder inom optimering användas. Annars måste aposterioririsken approximeras genom simulering från aposteriorifördelningen och sedan minimeras.

8.2 Bayesiansk hypotesprövning

Som vi har använt den hittills ger beslutsteorin ett verktyg för att utvärdera skattningar och identifiera ”optimala” skattningar. Vi antydde dock tidigare att handlingsrummet $\mathcal{A}$ inte nödvändigtvis behöver sammanfalla med parameterrummet. Det kan bestå av en godtycklig mängd handlingar a_1, a_2, a_3 osv. som vi kan ta beroende på vår

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

observation $X = x$. Vi har då vad som kallas en beslutsregel $W(x) \in \mathcal{A}$ som anger hur en viss observation leder till ett visst beslut eller handling. Det enda som krävas är att vi har någon förlustfunktion $L(\theta, a)$ som kvantifierar handlingens "korrekthet" givet ett visst parametervärde $\theta \in \Omega$.

8.2.1 Bayesbeslut

Givet en förlustfunktion och en apriorifördelning kan vi på samma sätt som innan definiera riskfunktionen $R(\theta, W)$ och bayesrisken $\mathrm{E}[R(\Theta, W)]$ för någon beslutsregel $W : \mathbb{R}^m \to \mathcal{A}$. När vi talar om allmänna beslutsregler $W(\boldsymbol{x})$ kallas den regel som minimerar bayesrisken för den *bayesianska beslutsregeln* eller *bayesbeslutet* (den kallas också för den *formella bayesregeln*, men då detta mest skapar förvirring undviker vi denna benämning). För att ta fram bayesbeslutet kan vi på samma sätt som innan minimera aposterioririsken

$$\mathrm{E}[L(\Theta, a) \mid \boldsymbol{X} = \boldsymbol{x}]$$

med avseende på $a = W(\boldsymbol{x})$ för varje $\boldsymbol{x} \in \mathbb{R}^m$. Genom att välja den handling a som minimerar denna risk kan vi då fatta det optimala beslutet (med avseende på förlustfunktionen och apriorifördelningen, så klart).

Exempel 8.5. En bonde funderar på att investera i ett litet vindkraftverk för att få ner sina elkostnader. Hon har två modeller att välja på: den mindre modellen a_1 som kostar 11 400 kr med radie 0.8 m och en större modell a_2 som kostar 67 900 kr med radie 1.5 m. Den teoretiska maxeffekten hos ett vindkraftverk ges av formeln

$$P = \frac{1}{2}\rho A v^3,$$

där P är effekten i watt, $\rho = 1.29$ kg/m^3 är luftdensiteten, A är arean (i m^2) som rotorbladen sveper över och v är vindhastigheten (i m/s). Hon räknar på ett genomsnittligt elpris på 3 kr/kWh (dvs. 0.003 kr/Wh) och vill maximera sin vinst under de kommande fem åren.

Den okända variabeln här (som måste skattas) är vindhastigheten v. Förlustfunktionen (förlust är negativ vinst) blir då för a_1 lika med

$$L(v, a_1) = 11400 - \frac{1}{2} \cdot 1.29 \cdot \pi \cdot 0.8^2 v^3 \cdot 24 \cdot 365 \cdot 5 \cdot 0.003 = 11400 - 170.5v^3$$

eftersom vi har $24 \cdot 365 \cdot 5$ timmar under fem år. På samma sätt har vi

$$L(v, a_2) = 67900 - \frac{1}{2} \cdot 1.29 \cdot \pi \cdot 1.5^2 v^3 \cdot 24 \cdot 365 \cdot 0.003 = 67900 - 600v^3.$$

Under ett par veckor antecknar hon vindhastigheten (i m/s) på platsen hon tänker placera vindkraftverket och får värdena

5.80 7.88 2.97 4.33 3.25 6.83 1.73

Om vi antar att vindhastigheterna är betingat oberoende och normalfördelade givet väntevärdet och variansen får vi att den marginella aposteriorifördelningen för väntevärdet ges av

$$V \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathrm{t}(6, 4.68, 4.98/7) = \mathrm{t}(6, 4.68, 0.71)$$

eftersom $\bar{x} = 4.68$ och $s^2 = 4.98$.

För att beräkna aposterioririsken exakt behöver vi beräkna det tredje momentet hos den generaliserade t-fördelningen, vilket är ganska krävande. Istället kan vi approximera aposterioririsken genom att generera ett stort antal (10000) utfall från $\mathrm{t}(6, 4.68, 0.71)$ och ta medelvärdena av $L(v, a_1)$ och $L(v, a_2)$. Vi får då

$$\mathrm{E}[L(V, a_1) \mid X = x] \approx -8480$$
$$\mathrm{E}[L(V, a_2) \mid X = x] \approx -2060.$$

Med tanke på vindförhållandena är det alltså bättre att köpa den mindre modellen. (Observera dock att med en längre tidshorisont skulle den större modellen vara bättre placerad.)

This document is available free of charge on studocu
Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

8.2.2 Hypoteser

En speciell (men vanligt förekommande) typ av handling är en hypotes. Med detta avser vi en undermängd Ω_0 av parameterrummet Ω som jämförs med en annan mängd Ω_1. Oftast har vi att $\Omega_1 = \Omega \setminus \Omega_0$ så att $\Omega_0 \cap \Omega_1 = \emptyset$ och $\Omega_0 \cup \Omega_1 = \Omega$. Vi har då två handlingar: a_0 "acceptera $H_0 : \boldsymbol{\theta} \in \Omega_0$" och a_1 "acceptera $H_1 : \boldsymbol{\theta} \in \Omega_1$".

En naturlig förlustfunktion i detta fall är det så kallade 0–1-felet:

$$L(\theta, a_0) = \begin{cases} 0, & \theta \in \Omega_0, \\ 1, & \theta \in \Omega_1, \end{cases} \qquad L(\theta, a_1) = \begin{cases} 1, & \theta \in \Omega_0, \\ 0, & \theta \in \Omega_1. \end{cases}$$

Med andra ord har vi en förlust på noll om vi gör "rätt" val och ett annars. Vi vill alltså undvika fel av första slaget (att vi accepterar H_1 när $\theta \in \Omega_0$) lika mycket som fel av andra slaget (att vi accepterar H_0 när $\theta \in \Omega_1$). I detta fall får vi aposterioririsken till

$$\begin{aligned} \mathrm{E}[L(\Theta, a_0) \mid X = x] &= \mathrm{P}(\Theta \in \Omega_1 \mid X = x) \\ \mathrm{E}[L(\Theta, a_1) \mid X = x] &= \mathrm{P}(\Theta \in \Omega_0 \mid X = x). \end{aligned}$$

Bayesbeslutet blir då a_0 om $\mathrm{P}(\Theta \in \Omega_0 \mid X = x) > \mathrm{P}(\Theta \in \Omega_1 \mid X = x)$ och a_1 annars. Vi väljer alltså den hypotes som har högst aposteriorisannolikhet.

Mer allmänt har vi det generaliserade 0–1-felet:

$$L(\theta, a_0) = \begin{cases} 0, & \theta \in \Omega_0, \\ c_{\mathrm{II}}, & \theta \in \Omega_1, \end{cases} \qquad L(\theta, a_1) = \begin{cases} c_{\mathrm{I}}, & \theta \in \Omega_0, \\ 0, & \theta \in \Omega_1, \end{cases}$$

där c_{I} motsvarar kostnaden för ett fel av första slaget (accepterar H_1 när vi borde acceptera H_0) och c_{II} motsvarar kostnaden för ett fel av andra slaget. I detta fall har vi

$$\begin{aligned} \mathrm{E}[L(\Theta, a_0) \mid X = x] &= c_{\mathrm{II}}\,\mathrm{P}(\Theta \in \Omega_1 \mid X = x) \\ \mathrm{E}[L(\Theta, a_1) \mid X = x] &= c_{\mathrm{I}}\,\mathrm{P}(\Theta \in \Omega_0 \mid X = x). \end{aligned}$$

Bayesbeslutet är då a_0 om

$$c_{\mathrm{II}}\,\mathrm{P}(\Theta \in \Omega_1 \mid X = x) < c_{\mathrm{I}}\,\mathrm{P}(\Theta \in \Omega_0 \mid X = x),$$

dvs. om

$$\frac{\mathrm{P}(\Theta \in \Omega_0 \mid X = x)}{\mathrm{P}(\Theta \in \Omega_1 \mid X = x)} > \frac{c_{\mathrm{II}}}{c_{\mathrm{I}}},$$

annars väljer vi a_1. Vänsterledet kallas här för aposterorioddset för H_0 mot H_1. Om detta värde är tillräckligt stort (där "tillräckligt" avgörs av de relativa kostnaderna c_{I} och c_{II}) så accepterar vi alltså H_0 – annars accepterar vi H_1.

Eftersom $\Omega_1 = \Omega \setminus \Omega_0$ kan vi förenkla bayesbeslutet ytterligare. Vi har $\mathrm{P}(\Theta \in \Omega_1 \mid X = x) = 1 - \mathrm{P}(\Theta \in \Omega_0 \mid X = x)$, vilket ger

$$\frac{\mathrm{P}(\Theta \in \Omega_0 \mid X = x)}{\mathrm{P}(\Theta \in \Omega_1 \mid X = x)} > \frac{c_{\mathrm{II}}}{c_{\mathrm{I}}}$$

om och endast om

$$\frac{\mathrm{P}(\Theta \in \Omega_0 \mid X = x)}{1 - \mathrm{P}(\Theta \in \Omega_0 \mid X = x)} > \frac{c_{\mathrm{II}}}{c_{\mathrm{I}}},$$

vilket i sin tur kan skrivas om som

$$\mathrm{P}(\Theta \in \Omega_0 \mid X = x) > \frac{c_{\mathrm{II}}}{c_{\mathrm{I}} + c_{\mathrm{II}}},$$

Observera att till skillnad från i frekventistisk hypotesprövning så behöver vi inte ange någon signifikansnivå α. Att acceptera H_0 eller H_1 beror helt och hållet på förlustfunktionens utseende, vilket i sin tur beror på tillämpningen.

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

8.2.3 Bayesfaktor

Ett annat sätt att göra hypotesprövning inom bayesiansk inferens är att jämföra hur oddsen ändras efter att datat observeras. Ovan hade vi aposteriorioddset. Vi kan också definiera apriorioddset

$$\frac{\mathrm{P}(\Theta \in \Omega_0)}{\mathrm{P}(\Theta \in \Omega_1)},$$

dvs. oddset enligt apriorifördelningen. Om vi tar kvoten mellan aposteriorioddset och apriorioddset får vi *bayesfaktorn*

$$\frac{\mathrm{P}(\Theta \in \Omega_0 \mid X = x) / \mathrm{P}(\Theta \in \Omega_1 \mid X = x)}{\mathrm{P}(\Theta \in \Omega_0) / \mathrm{P}(\Theta \in \Omega_1)} = \frac{\mathrm{P}(\Theta \in \Omega_0 \mid X = x)\, \mathrm{P}(\Theta \in \Omega_1)}{\mathrm{P}(\Theta \in \Omega_1 \mid X = x)\, \mathrm{P}(\Theta \in \Omega_0)}.$$

Tanken är att reducera inflytandet av apriorifördelningen och fokusera på hur observationerna påverkar förhållandet mellan hypoteserna. Apriorifördelningen har så klart en påverkan (förutom fallet med ”enkla hypoteser” $\Omega_0 = \{\theta_0\}$ och $\Omega_1 = \{\theta_1\}$), men *bayesfaktorn* gör denna påverkan mindre kännbar.

Exempel 8.6. Antag att vi mäter kvaliteten hos komjölk på en indexerad skala där 100 är medelvärdet över hela populationen. Vi har en apriorifördelning $\Theta \sim \mathrm{N}(100, 15^2)$ och en datafördelning $X \mid \Theta = \theta \sim \mathrm{N}(\theta, 10^2)$.

Nu vill vi testa om ett visst prov ligger under eller över medelvärdet. Vi har då $H_0 : \theta > 100$ och $H_1 : \theta \leq 100$. Eftersom apriorifördelningen är symmetrisk runt 100 har vi att apriorioddset är ett. Vi observerar $X = 115$ och aposteriorifördelningen är då $\mathrm{N}(110.39, 69.23)$. Detta ger aposteriorioddset:

$$\frac{\mathrm{P}(\Theta > 100 \mid X = 115)}{\mathrm{P}(\Theta \leq 100 \mid X = 115)} \approx \frac{0.894}{0.106} \approx 8.44.$$

Eftersom apriorioddset är lika med ett får vi också att bayesfaktorn är lika med 8.44. Detta är en ganska stor bayesfaktor och indikerar att hypotesen H_0 sannolikt stämmer.

8.2.4 Punkthypoteser för kontinuerliga parametrar

En sak som skiljer sig mellan bayesiansk och frekventistiskt hypotesprövning är hur man behandlar punkthypoteser, dvs. hypoteser på formen $H_0 : \theta = \theta_0$, för kontinuerliga parametrar. Inom det frekventistiska ramverket är det vanligt att specificera nollhypotesen $H_0 : \theta = \theta_0$ och jämföra denna med $H_1 : \theta \neq \theta_0$. För kontinuerliga variabler i bayesiansk inferens så fungerar inte detta eftersom vi alltid har $\mathrm{P}(\theta = \theta_0) = 0$ och $\mathrm{P}(\theta \neq \theta_0) = 1$, både i apriorifördelningen och aposteriorifördelningen.

Ett kontraargument är att punkthypoteser för kontinuerliga parametrar egentligen inte är rimliga för de flesta tillämpningar. Oftast så spelar det inte någon roll om ett parametervärde θ är exakt lika med θ_0. Det som är viktigt är om θ ligger inom ett visst intervall runt θ_0. Man kan då bilda hypotesen $H_0 : |\theta - \theta_0| \leq \epsilon$ för något $\epsilon > 0$. En sådan hypotes kan då ha positiv sannolikhet.

En annan möjlighet är att specificera en blandad apriorifördelning (blandad i och med att den är både diskret och kontinuerlig). I det fallet så antar vi att vi har två oberoende stokastiska variabler H och $Z \sim \mathrm{Ber}(p)$ för någon $p > 0$ samt ett värde θ_0 så att

$$\mathrm{P}(\Theta = \theta \mid Z = 1) = \begin{cases} 1, & \theta = \theta_0, \\ 0, & \text{annars.} \end{cases}$$

och

$$\mathrm{P}(\Theta \leq \theta \mid Z = 0) = F_H(z),$$

dvs. att $\Theta = \theta_0$ om $Z = 1$ och $\Theta = H$ om $Z = 0$. I detta fall har vi en positiv sannolikhet p att $\Theta = \theta_0$ och med sannolikhet $1 - p$ har den en annan (kontinuerlig) fördelning.

This document is available free of charge on studocu

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)

Nedladdat av Lucas Frykman (l.m.frykman@gmail.com)