Tommy Norberg
Matematisk statistik
Chalmers & GU
1 mars 2005

Introduktion till Bayesiansk uppdatering

Detta ska handla om hur man på ett vetenskapligt objektivt (= begåvat) sätt kan uppdatera kunskap, men ändå tillåta subjektiva värderingar av sin egen eller andras erfarenhet. Redskapet som vi ska lära oss använda är en generalisering av Bayes formel.

Ofta när man är intresserad av en viss storhet, θ säg, så är det så att antingen känner man storhetens exakta värde, t.ex $\theta = 0.35$, eller så känner man den med en felangivelse, t.ex $\theta = 0.35 \pm 0.15$. Felet kan här vara absolut i betydelsen att man säkert vet att $0.20 \leq \theta \leq 0.50$, men det kan också vara givet med en viss konfidens så som är fallet då det är uträknat med statistisk metodik.

I denna föreläsning ska vi lära oss ett annorlunda sätt att tänka och värdera kunskap om sannolikheter, men metodiken vi ska gå igenom kan tillämpas även på andra typer av intressanta storheter.

Vi ska tänka oss att vi har en reellvärd positiv funktion $p(\theta)$, definierad för $0 \leq \theta \leq 1$, som anger hur pass troligt varje tänkbart värde av sannolikheten θ är. Det enda kravet vi ställer på $p(\theta)$ är att

$$p(\theta) \geq 0 \quad \text{för alla} \quad \theta \in [0, 1]$$

Sådana funktioner kallas i statistiken för *trolighetsfunktioner* oftast bara *troligheter* (”likelihood” på engelska). Ibland beror de av en eller flera parametrar, t.ex α, β. Då betecknas troligheten $p(\theta|\alpha, \beta)$ och om vi t.ex vet att $\alpha = 3$ medan $\beta = 7$, så kanske vi skriver $p(\theta|\alpha = 3, \beta = 7)$ eller, om det inte kan missförstås, bara $p(\theta|3, 7)$

Om $p(\theta)$ är en trolighet och $q(\theta) = cp(\theta)$, där $c > 0$, så är $p(\theta)$ och $q(\theta)$ ekvivalenta i den betydelsen att de svarar mot exakt samma kunskap (eller information) om θ. Detta innebär att vi endast behöver specificera troligheter modulo en multiplikativ konstant.

Exempel 1 Antag att vi inte vet något alls om sannolikheten θ. Då kan det vara praktiskt att välja $p(\theta) = 1$ för $0 \leq \theta \leq 1$. □

Exempel 2 Antag att vi säkert vet att θ ligger mellan 0.2 och 0.7, samt att det troligaste värdet är 0.35. Även om det inte rekommenderas i denna föreläsning, skulle det kunna vara vettigt att välja $p(\theta)$ triangulär så att

$$p(\theta) = \begin{cases} \dfrac{\theta - 0.2}{0.15} & \text{då} \quad 0.2 \leq \theta \leq 0.35 \\ \dfrac{0.7 - \theta}{0.35} & \text{då} \quad 0.35 \leq \theta \leq 0.7 \end{cases}$$

och lika med 0 f.ö. $\square$

Definition 1 Vi ska kalla $p(\theta)$ för θ:s *à priori-trolighet*. (À priori är latin och betyder på förhand.) Om det skulle vara så att $\int_0^1 p(\theta)\, d\theta = 1$, så kan vi också kalla $p(\theta)$ för θ:s *à priori-täthet*.

En, såväl av praktiska som av teoretiska orsaker, ofta använd à prioritrolighet är

$$b(\theta|\alpha,\beta) \propto \theta^{\alpha-1}\,(1-\theta)^{\beta-1}$$

för lämpligt valda $\alpha, \beta > 0$. Tecknet $\propto$ betyder *proportionell mot*. Observera att proportionalitetskonstanten får bero av α och β, men ej av θ. Att skriva $b(\theta|\alpha,\beta) \propto \theta^{\alpha-1}\,(1-\theta)^{\beta-1}$ betyder att

$$b(\theta|\alpha,\beta) = c(\alpha,\beta)\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1}$$

Man brukar då säga att man har en beta(α,β)-*prior*.

När man sysslar med uppdatering av sannolikhetsskattningar har beta-troligheten, som vi ska se nedan, många värdefulla egenskaper. Medelst maximering (j.f.r övning 1) kan man visa att beta(α,β)-troligehetens maximum uppnås för sannolikheten

$$\widehat{\theta} = \frac{\alpha-1}{\alpha+\beta-2}$$

En förutsättning för att resultatet ska gälla är att både $\alpha, \beta > 1$.

Definition 2 Låt $p(\theta)$ vara en trolighet. Om det finns ett unikt $\widehat{\theta}$, sådant att

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\; p(\theta)$$

så säger man ofta att $\widehat{\theta}$ är troligehetens *mod*.

Övning 1 Visa att beta(α,β)-troligehetens mod är

$$\widehat{\theta} = \frac{\alpha-1}{\alpha+\beta-2}$$

om $\alpha, \beta > 1$.

Beta-troligheten kan normeras till en täthet. Vi låter

$$b(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1} \quad \text{för} \quad 0<\theta<1$$

Här är Γ den s k gammafunktionen, given av

$$\Gamma(z) = \int_0^\infty t^{z-1}\,e^{-t}\,dt \quad \text{för} \quad z>0$$

Vi har att

$$\int_0^1 b(\theta|\alpha,\beta)\,d\theta = 1$$

så med denna normering blir $b(\theta|\alpha,\beta)$ en täthet. Motsvarande fördelning kallas betafördelningen. I figur 1 är en beta$(5,2)$-täthet plottad. Du hittar fakta om betafördelningen i bl.a Råde & Westergren [6] och formelsamlingen [4]. Man kan visa (j.f.r övningarna 2 och 3 nedan) att beta(α,β)-fördelningens väntevärde och varians är

$$\mu_\theta = \frac{\alpha}{\alpha+\beta} \quad \text{och} \quad \sigma_\theta^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$$

och vi har redan sett att dess mod är

$$\widehat{\theta} = \frac{\alpha-1}{\alpha+\beta-2}$$

då $\alpha,\beta>1$. Moden i à prioritroligheten kallas ibland för *à priori-* eller *förhandsskattningen* av θ. Beta$(5,2)$-tätheten i figur 1 har moden $\widehat{\theta}=0.8$ och väntevärdet $\mu_\theta=0.714$.

Övning 2 Visa att beta(α,β)-täthetens väntevärde är

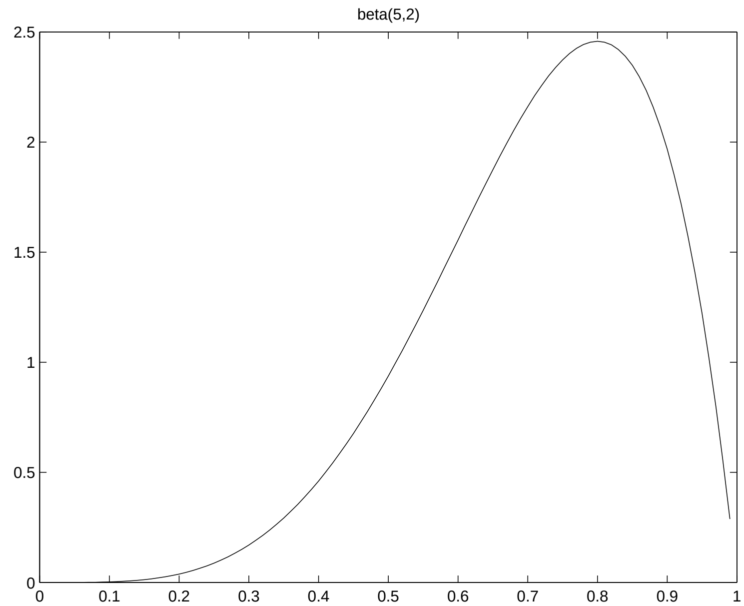
$$\mu_\theta = \frac{\alpha}{\alpha+\beta}$$

Ledning: $\Gamma(z+1)=z\Gamma(z)$ för $z>0$.

Övning 3 Visa att beta(α,β)-täthetens varians är

$$\sigma_\theta^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$$

Vi ska strax återkomma till hur man på basis av sin erfarenhet (kunskap, information, e.dyl) kan specificera parametrarna α,β i beta-priorn. Låt oss emellertid först titta på hur man medelst mätningar kan uppdatera en à prioritrolighet till en *à posteriori-dito* (à posteriori är latin och betyder i efterhand).



Figur 1: Beta$(5, 2)$-tätheten

Sats 1 (**Bayes formel**) *Låt $p(\theta)$ vara vår à prioritrolighet för θ. Antag att mätresultatet x, givet att θ är det sanna värdet på den okända sannolikheten, är fördelat enligt massfunktionen eller tätheten $p(x|\theta)$. Då gäller för à posterioritroligheten $p(\theta|x)$ att*

$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

Observera att normalt kan $p(\theta|x)$ normeras till en täthet. Sätt

$$c(x) = \left(\int_0^1 p(x|\theta)\, p(\theta)\, d\theta \right)^{-1}$$

Då gäller att

$$\int_0^1 c(x)\, p(x|\theta)\, p(\theta)\, d\theta = 1$$

Vi accepterar Sats 1 utan bevis.

Vi analogi med tidigare terminologi kallar vi moden i à posterioritroligheten, som är

$$\widehat{\theta}(x) = \operatorname*{argmax}_{\theta}\, p(\theta|x) = \operatorname*{argmax}_{\theta}\, p(x|\theta)\, p(\theta)$$

för *à posteriori*- eller efterhandsskattningen av θ. Notera att à posterioritroligheten är en funktion av utfallet x.

Exempel 3 Antag att vi som i exempel 1 inte har någon förutfattad uppfattning om θ och därför valt priorn $p(\theta) \propto 1$. Enligt Bayes formel ges à posterioritroligheten då av

$$p(\theta|x) \propto p(x|\theta)$$

À posterioriskattningen av θ är således då

$$\widehat{\theta}(x) = \underset{\theta}{\operatorname{argmax}}\ p(x|\theta)$$

I traditionell statistik kallar man denna à posterioriskattning för *trolighets-* (eller *ML-skattningen*) av θ. ML är en förkortning av engelskans "Maximum Likelihood". Läs själv om trolighetsskattning i avs 7.6 i Devore & Farnum [3]. Notera att ofta skriver man bara $\widehat{\theta}$ istället för $\widehat{\theta}(x)$. Att (à posteriori-, trolighets-) skattningen av θ beror av utfallet x är då underförstått. □

Övning 4 Låt θ vara sannolikheten att ett viss försök ska lyckas. Antag att du har gjort 10 oberoende försök, varav 3 st lyckades. Bestäm à posteriorifördelningen för θ om ingen förhandskunskap finns. Bestäm även à posterioriskattningen av θ.

Övning 5 (Fortsättning av övning 4.) Antag nu att du planerar att göra ytterligare m försök. Föreslå à prioritrolighet för θ.

Övning 6 (Fortsättning av övningarna 4, 5.) Antag att du erhöll $y = 5$ lyckade försök i dina $m = 12$ oberoende mätningar. Bestäm à posterioritrolighet och motsvarande skattning av θ.

Övningarna 4, 5 och 6 illustrerar en mycket användbar egenskap hos betafördelningen. Nämligen att om mätmodellen är $\mathrm{bin}(n, x)$ och à prioritätheten är $\mathrm{beta}(\alpha, \beta)$, så följer att à posterioritätheten är $\mathrm{beta}(\alpha + x, \beta + n - x)$. I Bayesiansk (statistisk) teori säger man att betafördelningen är konjugerad prior till binomialfördelningen. Detta skrives kortfattat

$$\boxed{\theta \sim \mathrm{beta}(\alpha, \beta),\ x|\theta \sim \mathrm{bin}(n, \theta) \quad \Rightarrow \quad \theta|x \sim \mathrm{beta}(\alpha + x, \beta + n - x)}$$

Vi förstår av det som sagts hittills att betafördelningen är en naturlig à priorifördelning då man vill öka sin kunskap om en sannolikhet θ genom att göra oberoende försök. Har man redan gjort n försök och erhållit x lyckade, väljes à priorifördelningen $\mathrm{beta}(x + 1, n - x + 1)$. Har man ingen förhandsinformation sättes här $n = x = 0$. À priorifördelning är då $\mathrm{beta}(1, 1) \propto 1$. Efter det att man observerat y lyckade försök bland de m nya försöken, övergår à priorifördelningen $\mathrm{beta}(x + 1, n - x + 1)$ i à posteriorifördelningen

$\text{beta}(x + y + 1, n + m - (x + y) + 1)$. Innan vi gjort den sista serien av m oberoende försök är trolighetsskattningen av θ lika med $\widehat{\theta} = x/n$ och efter det att vi gjort våra m nya försök och observerat y lyckade, är den uppdaterade trolighetsskattningen $\widehat{\theta} = (x + y)/(n + m)$. Detta är mycket tilltalande med tanke på att vi totalt gjort $n + m$ oberoende försök och observerat $x + y$ lyckade.

Övning 7 En ingenjör ska medelst provtagning avgöra hur stor sannolikheten θ är att ett visst gränsvärde överstigs. Hon gör därför 8 mätningar (som vi ska anta är oberoende upprepningar av försöket att bestämma en viss halt av någon förorening). I 3 av mätningarna överstegs gränsvärdet. Bestäm ingenjörens trolighetsskattning (med avseende på à posteriorifördelningen) av θ, om hon ej har någon förhandskunskap. Bestäm också à posteriorifördelningens standardavvikelse.

Övning 8 (Fortsättning på övning 7 ovan.) Anta nu att ingenjören tidigare gjort 4 mätningar och att gränsvärdet överstegs i en av dessa. Skriv upp lämplig à priorifördelning, och bestäm mod och standardavvikelse i à posteriorifördelningen. Jämför med fallet i övning 7 då hon inte hade någon förhandskunskap.

Ibland (kanske oftast) är det inte möjligt att exakt kvantifiera sin förhandskunskap. I nedanstående övningar indikeras hur man trots detta ändå kan tänka sig en à priorifördelning, som kan uppdateras med Bayes formel.

Övning 9 (Fortsättning på övning 7 ovan.) Anta nu istället att ingenjören erfarenhetsmässigt vet att $\theta \approx 0.25$ och att hon vill använda sig av denna erfarenhet i skattandet av θ. Hon bestämmer sig för att värdera sin erfarenhet lika mycket som den information hon erhåller från de 8 nya mätningarna. Föreslå à prioritäthet och bestäm à posterioriskattningen av θ. Om hennes erfarenhet är dubbelt så mycket värd som de 8 nya mätningarna, vilken à priorifördelning tycker du då är lämplig? Om hennes erfarenhet är värd hälften så mycket som de 8 nya mätningarna, vilken à priorifördelning tycker du då är lämplig?

Att som i övning 9, ovan, relatera värdet av ny kunskap till den befintliga verkar vara ett vettigt sätt att enkelt hitta en lämplig à priorifördelning i specifika tillämpningar. Man uttrycker då värdet av den nuvarande kunskapen genom att tillskriva den ett antal observationer. Det finns inget som säger att detta pseudo-antal observationer måste vara ett heltal.

Övning 10 Samma problemställning som i övning 9 ovan. Men nu anser ingenjören att hennes erfarenhet endast är värd ungefär 1/3 av den information hon får genom att utföra de 8 nya mätningarna. Föreslå ny à priorifördelning.

Övning 11 Samma problemställning som i övning 9 ovan. Men nu anser ingenjören att hennes tidigare erfarenhet svarar mot en à priorifördelning med standardavvikelsen $\sigma = 0.15$. Föreslå à priorifördelning.

Här följer ytterligare ett exempel på hur man kan gå till väga för att bestämma lämpliga värden på α, β då det är lämpligt med en beta-prior.

Exempel 4 Man önskar bebygga en f.d industritomt som har stått oanvänd ett antal år. Problemet är att det översta jordlagret är rejält förorenat av arsenik. I ett första planeringsskede vill man göra en preliminär uppskattning av den förorenade arean. Detta kan man göra genom att slumpa ut ett antal positioner helt godtyckligt, och räkna efter i hur många av dessa förorening detekteras. Låt oss anta att mätmetoden är utan fel, d.v.s att varken falska detekteringar eller falska icke-detekteringar förekommer. Då kommer antalet förorenade prover, x säg, att vara $\text{bin}(n, \theta)$-fördelat, där n är antalet utslumpade positioner och θ är proportionen förorenad mark. Vi har alltså mätmodellen $x|\theta \sim \text{bin}(n, \theta)$ och vi vet att det är praktiskt att då välja en beta-prior. Efter studium av tomtens historik och platsbesök kommer den lokala expertisen fram till att rimligen är ungefär 70% av arean förorenad. Av detta kan vi sluta oss till att

$$\widehat{\theta} = \frac{\alpha - 1}{\alpha + \beta - 2} \approx 0.70$$

eller

$$3\alpha \approx 7\beta - 4$$

Vi behöver ytterligare ett samband mellan α och β för att hitta lämpliga värden på båda parametrarna. Ett sådant skulle man kunna erhålla genom att uppmana experterna att försöka komma fram till en rimlig undre gräns θ_0 för den förorenade proportionen. Efter diskussion kom de fram till att det är mycket troligt att $\theta \geq 50\%$. Man ska kanske inte lite till 100% på experter, så vi ansätter därför att priorsannolikheten

$$P\{\theta \geq 0.5\} = \int_{0.5}^{1} b(\theta|\alpha, \beta)\, d\theta \approx 0.9$$

Med hjälp av numerisk programvara (t.ex Matlab) kan vi nu iterera fram att $\alpha = 8.77$ och $\beta = 4.33$ är lämpliga parametervärden. $\square$

Den som vill läsa mer om Bayesiansk statistik kan t.ex bläddra lite i Rice [5, kap 15], Bickel & Doksum [2] eller Berger [1]. Något exemplar av den förstnämnda finns ev att låna ut.

Referenser

[1] Berger, J A: Statistical Descision Theory: Foundations, Concepts and Methods, Springer-Verlag, New York, 1980.

[2] Bickel, Peter J & Kjell A Doksum: Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, San Fransisco, 1977.

[3] Devore, Jay & Nicholas Farnum: Applied Statistics for Engineers and Scientists. Duxbury Press, 1999.

[4] Norberg, Tommy: Formler och tabeller till matematisk statistik på universitet och tekniska högskolor. 2004. Ett ex kan laddas ner ifrån författarens hemsida `http:\\www.math.chalmers.se/~tommy`.

[5] Rice, John A: Mathematical Statistics and Data Analysis, second edition. Duxbury Press, 1995.

[6] Råde, Lennart & Bertil Westergren: Mathematics Handbook for Science and Engineering. Studentlitteratur, Lund, 1995.

Svar eller lösningar till övningarna

1) Sökt är

$$\underset{\theta}{\operatorname{argmax}}\ b(\theta|\alpha,\beta) = \underset{\theta}{\operatorname{argmax}}\ \theta^{\alpha-1}(1-\theta)^{\beta-1}$$
$$= \underset{\theta}{\operatorname{argmax}}\ ((\alpha-1)\ln\theta + (\beta-1)\ln(1-\theta))$$
$$= \frac{\alpha-1}{\alpha+\beta-2}$$

Den första likheten beror på att proportionalitetskonstanten bara påverkar maximats storlek, ej dess läge. Den andra likheten beror på att $b(\theta|\alpha,\beta)$ antar sitt maximum i samma punkt som $\ln b(\theta|\alpha,\beta)$ (logaritmfunktionen är ju strikt växande). Den tredje och sista likheten fås genom att sätta derivatan av $\ln b(\alpha,\beta)$ lika med noll och lösa ut θ. Till sist bör man försäkra sig om att lösningen är ett maximum.

2) Vi får

$$\mu_\theta = \int_0^1 \theta b(\theta|\alpha,\beta)\,d\theta = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^\alpha(1-\theta)^{\beta-1}\,d\theta$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\int_0^1 b(\theta|\alpha+1,\beta)\,d\theta = \frac{\alpha}{\alpha+\beta}$$

3) Använd tekniken från övning 2 till att visa att

$$\mu_{\theta^2} = \int_0^1 \theta^2 b(\theta|\alpha,\beta)\,d\theta$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}$$

och förtsätt så här:

$$\sigma_\theta^2 = \mu_{\theta^2} - (\mu_\theta)^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$$
$$= \frac{(\alpha+1)\alpha(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

4) Notera att mätresultatet x är bin(n,θ)-fördelat, där $n=10$. Motsvarande massfunktion är

$$p(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad \text{för} \quad x = 0,1,\ldots,n$$

Då vi inte har någon förhandskunskap väljes lämpligen priorn $p(\theta) \propto 1$ och Bayes formel ger för à posterioritroligheten att den satisfierar

$$p(\theta|x) \propto \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Observera att

$$p(\theta|x) \propto \theta^x (1-\theta)^{n-x}$$

Vi kan ju låta faktorn $\binom{n}{x}$ ingå i proportionalitetskonstanten $c(x)$. Vi ser att $p(\theta|x)$ är en beta$(x+1, n-x+1)$-trolighet. Motsvarande à posterioriskattning är $\widehat{\theta} = x/n$. Avslutningsvis noterar vi att då utfallet av x blev $x = 3$, så har vi alltså à posterioritroligheten beta$(4, 8)$ och à posterioriskattningen $\widehat{\theta} = 3/10 = 0.30$.

5) Naturligt är att välja beta$(x+1, n-x+1)$-troligheten till prior, där $n = 10$ och $x = 3$.

6) Priorn är $b(\theta|x+1, n-x+1)$ och mätresultatet y har massfunktionen

$$p(y|\theta) = \binom{m}{y} \theta^y (1-\theta)^{m-y}$$

Således gäller att à posterioritroligheten uppfyller

$$p(\theta|x, y) \propto \theta^{x+y} (1-\theta)^{n+m-(x+y)}$$

Detta är en beta$(x+y+1, n+m-(x+y)+1)$-trolighet. À posterioriskattningen är

$$\widehat{\theta} = \frac{x+y}{n+m}$$

Insättning av $n = 10$, $x = 3$ samt $m = 12$, $y = 5$ ger oss à posterioritroligheten beta$(9, 23)$ och-skattningen $\widehat{\theta} = 0.364$.

7) $\widehat{\theta} = 0.375$, $\sigma = 0.1477$

8) Beta$(2, 4)$, $\widehat{\theta} = 0.333$, $\sigma = 0.1237$

9) Beta$(3, 7)$, $\widehat{\theta} = 0.31$, beta$(5, 13)$, beta$(2, 4)$

10) Beta$(5/3, 3)$

11) Beta$(2.6, 5.9)$