Tommy Norberg
Matematisk statistik
Chalmers & GU
27 april 2004

Bayesiansk uppdatering av sannolikhets-skattningar

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. Vi kommer att generalisera Bayes formel till ett mycket användbart redskap, och använda detta redskap till att lära oss hur man lämpligen uppdaterar kunskap om sannolikheten för en händelse.

Ofta när man är intresserad av en viss storhet, θ säg, så är det så att antingen känner man storhetens exakta värde (t ex $\theta = 5.27$) eller så känner man den med en felangivelse (t ex $\theta = 5.27 \pm 1.18$). Felet kan här vara absolut i betydelsen att man säkert vet att $4.09 \leq \theta \leq 6.45$, 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 framförallt 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 $p(\theta)$ som anger hurpass troligt varje tänkbart värde av θ är. Vi ska till att börja med låta denna funktion vara definierad på hela reella tallinjen, alltså för $-\infty < \theta < \infty$, men låta $p(\theta) > 0$ bara för sådana värden som vår storhet θ verkligen kan anta. I fallet då θ är en sannolikhet gäller att $p(\theta) > 0$ bara då $0 \leq \theta \leq 1$.

Exempel 1 Låt θ vara en sannolikhet som vi inte vet något om. Då kan det vara praktiskt att välja $p(\theta) = 1$ för $0 \leq \theta \leq 1$. (Observera att när vi skriver så här är det underförstått att $p(\theta) = 0$ för alla andra θ.)

Exempel 2 Låt θ vara en sannolikhet som vi säkert vet ligger mellan 0.2 och 0.7. Vi vet också att det troligaste värdet är 0.35. Då skulle det kunna

vara vettigt att välja $p(\theta)$ triangulär så att

$$p(\theta) = \begin{cases} 0 & \text{då} \quad 0 \le \theta \le 0.2 \\ \dfrac{\theta - 0.2}{0.15} & 0.2 \le \theta \le 0.35 \\ \dfrac{0.7 - \theta}{0.35} & 0.35 \le \theta \le 0.7 \\ 0 & 0.7 \le \theta \le 1 \end{cases}$$

Exempel 3 Låt θ vara en stokastisk variabels väntevärde. Antag t ex att vi studerar en slumpmässig tid T som vi vet är exponentialfördelad med väntevärde θ. Vi tror att $\theta \approx 10$. Ett av teoretiska skäl (som vi inte går närmare inpå) är ett lämpligt principiellt utseende av $p(\theta)$ fölande

$$p(\theta) = \frac{1}{\theta^n} e^{-s/\theta} \quad \text{för} \quad \theta > 0$$

där $n > 0$ och $s > 0$. Observera att $p(\theta)$ antar sitt maximala värde då $\theta = s/n$. Vi låter därför $s = 10n$. Då

$$p(\theta) = \frac{1}{\theta^n} e^{-10n/\theta} \quad \text{för} \quad \theta > 0$$

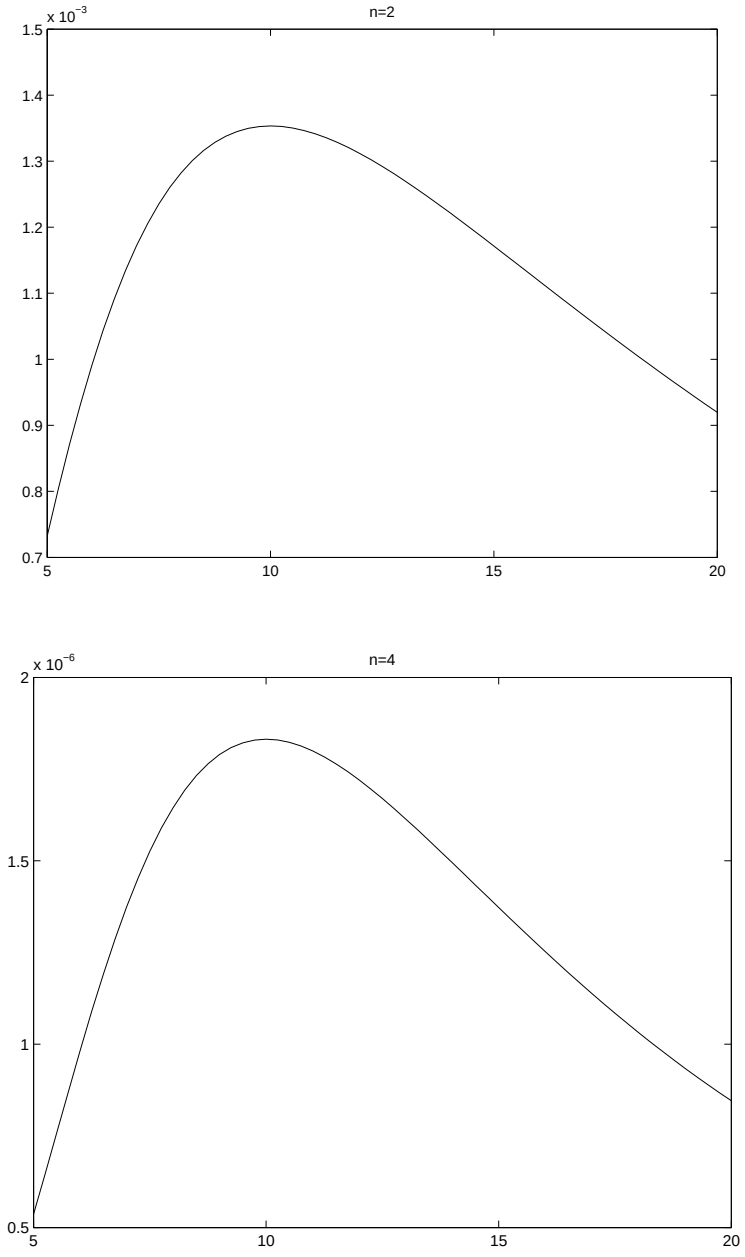
Det är nu lätt att övertyga sig om att parametern n på något sätt svarar mot säkerheten i vår tro att $\theta \approx 10$, för att ju större n är desto mer koncentrerad är $p(\theta)$ runt maximat vid $\theta = 10$. Se figur 1. Även här kan man tänka sig att modellera total okunskap om θ med en konstant funktion. T ex,

$$p(\theta) = 1 \quad \text{för} \quad \theta > 0$$

vilket svarar mot fallet $n = 0$.

Exempel 4 θ är två-dimensionell och lika med (μ, σ), där μ är väntevärdet och σ standardavvikelsen av en normalfördelad stokastisk variabel X.

De två funktionerna i figur 1 antyder att det kan vara svårt att jämföra två eller flera föreslagna $p(\theta)$-funktioner i ett konkret fall. Därför är det vanligt att man normerar så att totala ytan under kurvan blir lika med ett. (Förutsatt att det går.) Gör man så blir $p(\theta)$ en (sannolikhets-) täthet.



Figur 1: Plot av funktionen $p(\theta) = (1/\theta^n)\exp(-10n/\theta)$ för två olika värden på n (se exempel 3). I den översta är $n = 2$ medan $n = 4$ i den understa. Observera att y-axeln är så olika skalad i de två plottarna att det är svårt att visa båda funktionerna i samma plot

Definition 1 Vi ska kalla $p(\theta)$ för θ:s *à priori-täthet* eller *à priori-trolighet.* (À priori är latin och betyder på förhand.)

Vi tänker oss nu att vi gör ett eller flera oberoende slumpmässiga försök där vi observerar en stokastisk variabel X, vars massfunktion eller täthet beror av storheten θ. Vi ska låta $p(x|\theta)$ beteckna denna variabels mass- eller täthetsfunktion.

Exempel 5 Låt θ vara sannolikheten att en viss händelse A inträffar i ett försök och antag att vi gör n oberoende upprepningar av försöket. Låt X vara antalet lyckade försök. Då är $X|\theta$ binomialfördelad med parametrar n och θ. Motsvarande massfunktion är

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

Exempel 6 Låt θ vara väntevärdet av en exponentialfördelad tid och vi gör n oberoende mätningar av denna tid. Låt $x = x_1, \ldots, x_n$ beteckna de erhållna resultaten. Den täthet som beskriver detta experiment är

$$p(x|\theta) = \prod_i \left(\frac{1}{\theta} \, e^{-x_i/\theta} \right) = \frac{1}{\theta^n} \, e^{-\sum_i x_i/\theta} = \frac{1}{\theta^n} \, e^{-n\bar{x}/\theta}$$

Observera att alla $x_i > 0$, så medelvärdet $\bar{x} = \sum_i x_i/n > 0$.

Funktionen $p(x|\theta)$ är alltså bestämd av försöket man gör för att erhålla mer information om storheten θ. Man kan säga att den beskriver *mät- eller observationsmodellen.* Bayes formel, som vi känner från den grundläggande sannolikhetsteorin, kan användas för att motivera sats 1 nedan som talar om hur vi med hjälp av försöksresultatet ska uppdatera vår à priori-täthet $p(\theta)$. Tätheten som beskriver vår uppdaterade kunskap ska vi beteckna $p(\theta|x)$.

Definition 2 Vi ska kalla $p(\theta|x)$ för *θ:s à posteriori-täthet.* (À posteriori är latin och betyder i efterhand.)

À posterioritätheten beskriver alltså vår uppdaterade kunskap om θ som vi har efter det att vi gjort och analyserat försöket.

Sats 1 (Bayes sats) *À posterioritätheten $p(\theta|x)$ är proportionell mot produkten av modell-tätheten (eller massfunktionen) $p(x|\theta)$ och à priori-tätheten $p(\theta)$, d v s*

$$p(\theta|x) \propto p(x|\theta) \, p(\theta) \tag{1}$$

Tecknet $\propto$ betyder proportionell mot. Den som siktar på en djupare förståelse av Bayes sats kan med fördel ta sig an följande utvigdning av Bayes formel så som vi känner den från den grundläggande sannolikhsläran till ett resultat som gäller för diskreta stokastiska variabler. Andra hoppar till stycket som börjar omedelbart efter QED på sidan 7.

Vi börjar med att formulera Bayes formel något generellare än vad Devore & Farnum [3] gör i övning 13 på sidan 200. Låt $B_1, \ldots, B_n$ vara en partition av utfallsrummet och låt A vara en observerad händelse. Att $B_1, \ldots, B_n$ är en partition av utfallsrummet betyder att exakt en av dem alltid inträffar. Bayes formel lyder

$$P(B_j|A) = \frac{P(A|B_j)\,P(B_j)}{\sum_{j=1}^{n} P(A|B_j)\,P(B_j)} \tag{2}$$

Tanken i Bayes formel är den att efter det att vi observerat A vill vi räkna fram hur troliga de olika händelserna B_j är. Vill vi t ex gissa vilken av händelserna $B_1, \ldots, B_n$ som inträffat, ligger det nära till hands att gissa på den med högst betingad sannolikhet, alltså den som maximerar $P(B_j|A)$. Det kan vara värt att notera att för att få reda på vilken av sannolikheterna som är störst, räcker det att beräkna samtliga $P(A|B_j)\,P(B_j)$. Nämnaren i Bayes formel är ju densamma för alla fallen.

Vi ska nu formulera (2) med stokastiska variabler. Vi tänker oss att vi istället för att observera händelsen A, observerar utfallet x av en stokastisk variabel X. Vi identifierar alltså A med $X = x$. Låt dessutom Y vara den stokastiska variabel som antar värdet j precis då B_j infaller. Genom att observera vilket värde Y antar får vi då reda på exakt vilken av händelserna B_j som inträffar.

Bayes formel lyder med dessa omformuleringar

$$P(Y = j|X = x) = \frac{P(X = x|Y = j)\,P(Y = j)}{\sum_{j=1}^{n} P(X = x|Y = j)\,P(Y = j)} \tag{3}$$

Notera att faktorn

$$\frac{1}{\sum_{j=1}^{n} P(X = x|Y = j)\,P(Y = j)} = \frac{1}{P(X = x)}$$

fungerar som proportionalitetskonstant i uttrycket

$$P(Y = j|X = x) \propto P(X = x|Y = j)\, P(Y = j)$$

Tecknet $\propto$ ska utläsas proportionell mot. Om vi ser j som ett typiskt utfall av Y och x som ett dito av X, kan ovanstående skrivas med hjälp av betingade och obetingade massfunktioner

$$p(j|x) \propto p(x|j)\, p(j) \tag{4}$$

Här tänker vi oss att

$$p(j) = P(Y = j)$$
$$p(x|j) = P(X = x|Y = j)$$

och att

$$p(j|x) = P(Y = j|X = x)$$

Observera att $p(j|x) \propto p(x|j)\, p(j)$ betyder

$$p(j|x) = C(x)\, p(x|j)\, p(j)$$

där $C(x)$ fås ur villkoret

$$\sum_{j=1}^{n} p(j|x) = 1$$

som leder till att $C(x)$ ges av

$$C(x) = 1/\textstyle\sum_{j=1}^{n} p(x|j)\, p(j)$$

och vi ser att formel (4) egentligen säger oss att

$$p(j|x) = \frac{p(x|j)\, p(j)}{\sum_{j=1}^{n} p(x|j)\, p(j)}$$

(Jämför med formel (3).)

När vi till sist tar steget tillformel (1) tänker vi på Y som något naturligt fenomen: en sannolikhet, en fysikalisk konstant, hållfastheten hos en konstruktion (t ex bro), etc, och byter beteckning från j till det mer etablerade θ. Värdet på storheten/parametern θ är ej känt. Vi fortsätter att tänka på x

som det vi observerar, och erhåller då uttrycket (1), *viz*

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

som är det generella sättet att skriva *Bayes formel*. Observera att parametern θ i de allra flesta fall är en kontinuerlig storhet och att ovanstående argumentering som leder fram till sats 1 ej är ett bevis av Bayes formel i det kontinuerliga fallet. QED

I den situation vi ska fokusera på är $p(x|\theta)$ en massfunktion och θ är en sannolikhet som i princip kan anta alla värden mellan 0 och 1. Vi ska normera så att $p(\theta)$ och $p(\theta|x)$ är tätheter. Bayes formel (1) är då ett kortfattat sätt att skriva

$$p(\theta|x) = C(x)\, p(x|\theta)\, p(\theta)$$

där proportionalitetskonstanten $C(x)$ bestäms av att

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

Vi får då att

$$C(x) = \frac{1}{\int_0^1 p(x|\theta)\, p(\theta)\, d\theta}$$

Så

$$p(\theta|x) = \frac{p(x|\theta)\, p(\theta)}{\int_0^1 p(x|\theta)\, p(\theta)\, d\theta}$$

Övning 1 Antag att

$$p(x|\theta) \propto \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(a) För vilka värden på θ är detta en massfunktion i x?
(b) För sådana θ, hur ser det exakta uttrycket för $p(x|\theta)$ ut?

Övning 2 Antag att

$$p(x|\theta) \propto e^{-x/\theta}, \quad x > 0$$

(a) För vilka värden på θ är detta en sannolikhetstäthet i x?
(b) För sådana θ, hur ser det exakta uttrycket för $p(x|\theta)$ ut?

Övning 3 Antag att

$$p(\theta|x) \propto \theta e^{-\theta x}, \quad \theta > 0$$

där observationen är $x > 0$. Visa att

$$p(\theta|x) = x^2 \theta e^{-\theta x}, \quad \theta > 0$$

Tanken med Bayes formel (1) är att vi är intresserade av någon storhet θ, om vilken vi har förhandsinformation i form av en sannolikhetstäthet $p(\theta)$ som helt enkelt talar om för oss hur troliga vi tror att olika värden på θ är. Vi uttrycker detta genom att säga att $p(\theta)$ är θ:s à priori- eller förhandstäthet. Sannolikhetstätheten $p(x|\theta)$ definierar mät- eller observations-modellen. Den beskriver hur sannolika de olika utfallen x är givet att just θ är det verkliga parametervärdet. Sannolikhetstätheten $p(\theta|x)$ talar om för oss hur troliga de olika värdena av θ är efter det att vi observerat just x. Den definierar θ:s à posteriori- eller efterhandsfördelning.

Ofta är dessa sannolikhetstätheter endast kända så när som på någon konstant som alltid kan bestämmas medelst normering. Anta t ex att θ är en sannolikhet med à posteriori-fördelning

$$p(\theta|x, y) \propto \theta^x (1-\theta)^y$$

(Här tänker vi oss att vi observerat heltal $x, y \geq 0$ och medelst multiplikation av en modell-massfunktion $p(x, y|\theta)$ med en à priori-täthet $p(\theta)$ erhållit just $\theta^x(1-\theta)^y$. Eventuellt har vi kastat faktorer som ej innehåller θ.) Då gäller ju, enligt vad som sagts ovan, att

$$p(\theta|x, y) = C(x, y)\, \theta^x (1-\theta)^y$$

och ur kravet

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

fås att

$$C(x, y) = \frac{1}{\int_0^1 \theta^x (1-\theta)^y\, d\theta}$$

I matematiken visas att

$$C(x, y) = \frac{\Gamma(x+y+2)}{\Gamma(x+1)\Gamma(y+1)} = \frac{(x+y+1)!}{x!\, y!}$$

där Γ är den s k gammafunktionen, given av

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

Jämför med t ex Råde & Westergren [5].

I resten av denna text ska vi låta θ vara sannolikheten att ett viss försök ska lyckas. Det kan handla om att sannolikheten att detektera en viss förorening, det kan handla om att sannolikheten att detektera att en patient har en viss sjukdom, t ex bröstcancer, det kan handla om sannolikheten att en viss typ av stänger går sönder i ett draghållfasthetsprov, m m. Vi ska göra n oberoende upprepningar av försöket i syfte att få fram information om θ. Anta först att vi inte har någon à priori-kunskap om θ utöver $0 < \theta < 1$. Det kan då vara naturligt att som à priori-fördelning välja den likformiga, som ju har konstant täthet, så vi låter

$$p(\theta) = 1 \quad \text{för} \quad 0 < \theta < 1$$

Låt X vara antalet lyckade försök. Då är $X \sim \text{bin}(n, \theta)$. Motsvarande massfunktion är

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

Bayes formel ger nu att à posteriori-kunskapen representeras utav sannolikhetstätheten

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

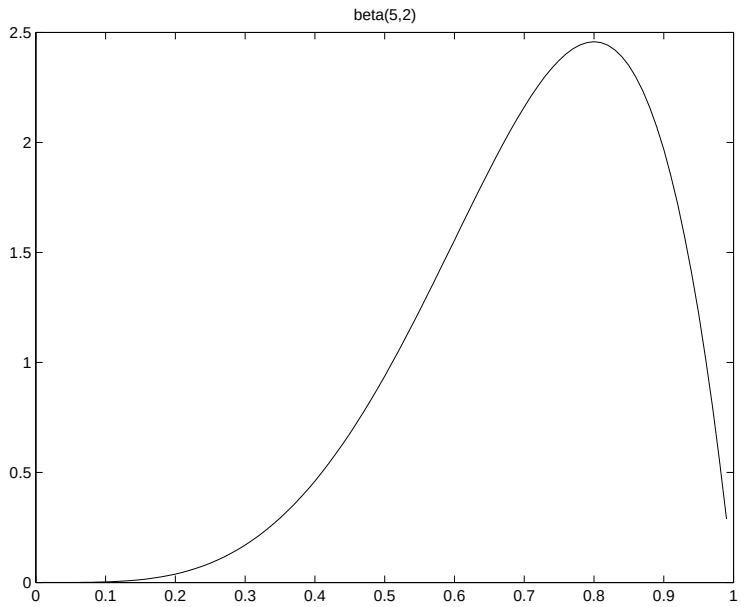
Observera att faktorer som ej innehåller θ behöver ej skrivas ut explicit, eftersom de då döljs i proportionalitetskonstanten. Att $p(\theta|x)$ är en beta$(x+1, n-x+1)$-täthet ser vi i t ex Råde & Westergren [5].

Ur [5] lär vi oss nu att en beta(p, q)-fördelning definitionsmässigt har tätheten

$$f(\theta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\, \theta^{p-1}\, (1-\theta)^{q-1} \quad \text{för} \quad 0 < \theta < 1$$

Ett exempel på hur denna kan se ut visas i figur 2. I [5] får vi också reda på att beta(p, q)-fördelningens väntevärde är

$$E[\theta] = \frac{p}{p+q}$$



Figur 2: beta$(5, 2)$-fördelningens täthet

och att dess varians är

$$\mathrm{Var}[\theta] = \frac{pq}{(p+q)^2(p+q+1)}$$

Medelst maximering av beta(p, q)-tätheten ser vi dessutom att dess troligaste utfall är

$$\widehat{\theta} = \frac{p-1}{p+q-2}$$

Detta värde kallas ibland för beta-fördelningens *mod*.

Övning 4 Visa att beta(p, q)-fördelningens mod är $\widehat{\theta} = (p-1)/(p+q-2)$.

Övning 5 Bestäm väntevärde, standardavvikelse och mod för fördelningen i figur 2.

Om man vill gissa ett värde på parametern θ är det naturligt i detta sammanhang att använda à posteriori-fördelningens mod, eftersom den anger det troligaste värdet. Ett inte lika naturligt val är à posteriori-fördelningens väntevärde. Standardavvikelsen i à posteriori-fördelningen är ett bra mått

på dessa skattningars osäkerhet. Ju mindre standardavvikelsen är, desto mer koncentrerad är tätheten runt moden och desto bättre är skattningen. Beteckningen $\widehat{\theta}$ för moden är inte standard. Vi använder den här för att i mer "normala" statistiska sammanhang är $\widehat{\theta}$ den s k trolighetsskattningen av θ.

Det är nämligen så att moden i à posteriori-fördelningen, som vi beräknar med den konstanta priorn $p(\theta) \propto 1$, som vi tar till då vi inte har någon tidigare information, i icke Bayesianska sammanhang kallas för trolighetsskattningen (trolighet är en översättning av engelskans likelihood). I icke-Bayesiansk statistik betraktas ofta modell-tätheten $p(x|\theta)$ som en funktion av θ för varje fixt utfall x. Denna funktion kallas då *trolighetsfunktionen* och *trolighetsskattningens* värde för utfallet x är

$$\widehat{\theta}(x) = \arg\max_{\theta} p(x|\theta)$$

Läs själv om trolighetsskattning i avs 7.6 i Devore & Farnum [3]

Övning 6 Låt θ vara sannolikheten att ett viss försök ska lyckas. Antag att du har gjort 10 oberoende försök, varav 3 st lyckades. Bestäm à posteriori-fördelningen om ingen förhandskunskap finns.

Övning 7 I diskussionen ovan erhöll vi à posteriori-tätheten $p(\theta|x) \propto \theta^x\,(1-\theta)^{n-x}$. Detta är en beta$(x+1, n-x+1)$-fördelning. Härled ur detta faktum formler för väntevärde μ, varians σ^2, standardavvikelse σ och mod $\widehat{\theta}$. Räkna sedan ut dessa storheter med värdena från övning 6 insatta.

Anta nu att vi vid ett senare tillfälle skall göra ytterligare m försök för att få mer information om sannolikheten θ. Vår à priori-fördelning nu är à posteriori-fördelningen från föregående mättillfälle:

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

Den representerar ju den kunskap om θ som vi tillägnat oss hittills. Vi observerar nu $Y \sim \text{bin}(m, \theta)$. Motsvarande massfunktion är

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

Medelst multiplikation av modelltätheten $p(y|\theta)$ med à priori-tätheten $p(\theta|x)$ (jämför med Bayes formel) erhåller vi den nya eller uppdaterade à posteriori-

tätheten

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

Ett synnerligen naturligt resultat med tanke på att man gjort totalt $n+m$ mätningar, varav $x+y$ lyckades. Vi har ju att $X+Y$ är binomialfördelad med parametrar $n+m$ och θ.

Övning 8 (Fortsättning på övningarna 6 och 7.) Anta att man vid det andra mättillfället gjorde 5 oberoende försök, varav 2 lyckades. Skriv upp modell, en lämplig à priori-fördelning och bestäm à posteriori-fördelningen.

Övning 9 Bestäm väntevärde, standardavvikelse och mod för à posteriori-fördelningen i övning 8.

Vi förstår av det som sagts hittills att beta-fördelningen är en naturlig à priori-fördelning då man vill öka sin kunskapsmängd om sannolikheten θ för en viss händelse, om mätmodellen är binomial. Har man redan gjort n försök och erhållit x lyckade, väljes à priori-fördelningen beta$(x+1, n-x+1)$. Har man ingen förhandsinformation sättes $n = x = 0$. À priori-fördelning är då beta$(1,1)$. Kolla själv att motsvarande täthet är konstant. Efter det att man observerat y lyckade försök bland de m nya försöken, övergår à priori-fördelningen beta$(x+1, n-x+1)$ i à posteriori-fördelningen beta$(x+y+1, n+m-(x+y)+1)$. Innan vi gjort våra nya försök är trolighetsskattningen av θ lika med x/n och efter det att vi gjort våra m nya försök och observerat y lyckade, är trolighetsskattningen $(x+y)/(n+m)$. Detta är mycket tilltalande med tanke på att vi totalt gjort $n+m$ försök och observerat $x+y$ lyckade.

Övning 10 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 (m a p à posteriori-fördelningen) av sannolikheten θ, om hon ej har någon förhandskunskap. Bestäm också à posteriori-fördelningens standardavvikelse.

Övning 11 (Fortsättning på övning 10 ovan.) Anta nu att ingenjören tidigare gjort 4 mätningar och att gränsvärdet överstegs i en av dessa. Skriv upp lämplig à priori-fördelning, och bestäm mod och standardavvikelse i

à posteriori-fördelningen. Jämför med fallet i övning 10 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 priori-fördelning, som kan uppdateras med Bayes formel.

Övning 12 (Fortsättning på övning 10 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 priori-täthet och bestäm trolighetsskattningen av θ m a p à posteriori-fördelningen. Om hennes erfarenhet är dubbelt så mycket värd som de 8 nya mätningarna, vilken à priori-fördelning tycker du då är lämplig? Om hennes erfarenhet är värd hälften så mycket som de 8 nya mätningarna, vilken à priori-fördelning tycker du då är lämplig?

Att som i övning 12, ovan, relatera värdet av ny kunskap till den befintliga verkar vara ett vettigt sätt att enkelt hitta en lämplig à priori-fördelning 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 13 Samma problemställning som i övning 12 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 priori-fördelning.

Övning 14 Samma problemställning som i övning 12 ovan. Men nu anser ingenjören att hennes tidigare erfarenhet svarar mot en à priori-fördelning med standardavvikelsen $\sigma = 0.15$. Föreslå à priori-fördelning.

Den som trängtar efter mer kunskap om Bayesiansk uppdatering av kunskap kan t ex bläddra lite i Rice [4, kap 15], Bickel & Doksum [2] eller Berger [1]. Ett och annat ex av den förstnämnda finns att låna ut.

Facit till övningarna

1) (a) $\theta > 0$ (b) $p(x|\theta) = e^{-\theta}\theta^x/x!,\ x = 0, 1, 2, \ldots$
2) (a) $\theta > 0$ (b) $p(x|\theta) = e^{-x/\theta}/\theta,\ x > 0$

3) Ansätt $p(\theta|x) = C(x)\,\theta e^{-\theta x}$ och notera att $C(x)$ fås ur $1 = \int_0^\infty p(\theta|x)\,d\theta = C(x)\int_0^\infty \theta e^{-\theta x}\,d\theta = C(x)/x^2$, så $C(x) = x^2$, v s v.
4) Sökt är

$$\arg\max_\theta f(\theta) = \arg\max_\theta \theta^{p-1}\,(1-\theta)^{q-1}$$

$$= \arg\max_\theta \left((p-1)\ln\theta + (q-1)\ln(1-\theta)\right) = (p-1)/(p+q-2)$$

Den första likheten beror på att proportionalitetskonstanten bara påverkar maximats storlek, ej dess läge. Den andra likheten beror på att $f(\theta)$ antar sitt maximum i samma punkt som $\ln f(\theta)$ (logaritmfunktionen är ju strikt växande). Den tredje och sista likheten fås genom att sätta derivatan av $\ln f(\theta)$ lika med noll och lösa ut θ. Då ser man tydligt att proportionalitetskonstanten ej har med maximats läge att göra. Slutligen bör man försäkra sig om att lösningen är ett maximum och inte en minimipunkt.
5) $\mu = 5/7 \approx 0.714$, $\sigma \approx 0.1597$, $\widehat{\theta} = 4/5 = 0.8$
6) beta$(4, 8)$
7) $\mu = (x+1)/(n+2) \approx 0.33$, $\sigma^2 = (x+1)(n-x+1)/((n+2)^2(n+3)) \approx 0.0171 \;\Rightarrow\; \sigma \approx 0.131$, $\widehat{\theta} = x/n = 0.3$
8) Modellen $y \sim \text{bin}(5, \theta)$, à priori-fördelningen beta$(4, 8)$ och observationen $y = 2$, ger à posteriori-fördelningen beta$(6, 11)$
9) $\mu \approx 0.35$, $\sigma \approx 0.113$, $\widehat{\theta} \approx 0.33$
10) $\widehat{\theta} = 3/8 = 0.375$, $\sigma \approx 0.1477$
11) beta$(2, 4)$, $\widehat{\theta} = 4/12 = 1/3 \approx 0.333$, $\sigma \approx 0.1237$
12) beta$(3, 7)$ ger trolighetsskattningen $\widehat{\theta} \approx 0.31$ m a p à posteriori-fördelningen, beta$(5, 13)$, beta$(2, 4)$
13) beta$(5/3, 3)$
14) beta$(2.6, 5.9)$

Tack till

Lars Rosén, Olle Nerman och Rossitza Dodunekova, som med konstruktiva kommentarer hjälpt mig att utforma dessa anteckningar om Bayesianska sannolikhetsskattningar.

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

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