

11 Bayesiansk tillförlitlighetsteori

Klassisk statistisk inferens som vi lär ut på våra statistikkurser är frekvensbaserad eller frekventistisk. Anta som exempel att du vill skatta en sannolikhet p för en händelse A i ett försök. Det normala är att du då gör n st oberoende upprepningar av försöket och räknar efter hur många gånger A inträffar. Kalla detta antal för f . Då skattar vi p med den väntevärdesriktiga skattningen

$$\hat{p} = \frac{f}{n}$$

och använder normalapproximation för att bestämma $(1 - \alpha)$ -konfidensgränser. Vi kan då skriva

$$p = \hat{p} \pm t_{\alpha/2}(n-1) \sqrt{\hat{p}(1-\hat{p})/(n-1)}$$

förutsatt att antalet försök n ej är för få. Ett alternativ, som kan tas till då normalapproximationen inte är bra, är att beräkna exakta konfidensgränser m.h.a binomialfördelningen.

I Bayesiansk statistik försöker man utnyttja ev redan erhållen kunskap om sannolikheten p . Antag t.ex att man är tämligen säker på att $p \approx 0.8$. Detta kan man konkretisera matematiskt genom att ange en *à priori-täthet*¹ eller -trolighet för sannolikheten p . Detta är en funktion $\pi(p)$ definierad för $p \in [0, 1]$. Ett naturligt val av *à priori-täthet* är

$$\pi(p) = Cp^a(1-p)^b \quad (1)$$

där konstanten C ges av $\int_0^1 \pi(p) dp = 1$. Men observera att värdet av C inte det minsta påverkar slutsatserna i analysen som följer.² Därför brukar man inte bry sig om C och bara ange att $\pi(p)$ är proportionell mot $p^a(1-p)^b$, vilket skrives

$$\pi(p) \propto p^a(1-p)^b$$

parametrarna a, b väljs så att formen av $\pi(p)$ reflekterar den förhandskunskap man har om p . Denna form bestäms enbart av a, b . Vi noterar först att om man som ovan tror att $p \approx 0.8$, så bör man välja a, b så att

$$\frac{a}{a+b} \approx 0.8 \quad \Leftrightarrow \quad a \approx 4b$$

eftersom $\pi(p)$ antar sitt största värde då $p = a/(a+b)$. Vi noterar vidare att ju större summan $a+b$ är, desto mer koncentrerad blir $\pi(p)$ runt maximat. Känner man sig väldigt säker på att $p \approx 0.8$, ska man därför välja $a+b$ stort och, omvänt, känner man sig inte speciellt säker, så väljer man ett mindre värde på $a+b$.

Anta nu att vi har observerat att händelsen A inträffat f gånger i n oberoende försök. Vad som då behöver göras är att definiera en regel som omvandlar *à priori-tätheten* till en *à posteriori-dito*³ betecknad $\pi(p|n, f)$. Det finns goda skäl att använda följande

¹à priori är latin och betyder på förhand

²Åtminstone om man gör analysen pragmatiskt så som undertecknad förespråkar

³à posteriori är latin och betyder i efterhand

regel:

$$\pi(p | n, f) \propto \pi(f | n, p) \pi(p) \quad (2)$$

där $\pi(f | n, p)$ är f 's täthet, alltså den vanliga binomialtätheten,

$$\pi(f | n, p) = \binom{n}{f} p^f (1-p)^{n-f}$$

Uppdateringsregeln (2), som kallas för *Bayes sats*, är en generalisering av Bayes formel som går igenom i de flesta grundkurser i matematisk statistik. Observera att proportionaliteten i (2) är m.a.p p . Således gäller att

$$\pi(p | n, f) = C \pi(f | n, p) \pi(p)$$

där C ges utav att

$$\int_0^1 C \pi(f | n, p) \pi(p) dp = 1$$

Vi ser att

$$C = 1 / \int_0^1 \pi(f | n, p) \pi(p) dp$$

Vi utför multiplikationen i (2), kastar faktorer som saknar betydelse, och erhåller

$$\pi(f | n, p) \propto p^{a+f} (1-p)^{b+n-f}$$

Nu inses varför det är naturligt att välja å prioritätheten i (1). Att man får tillbaka samma typ av täthet efter uppdatering med Bayes sats brukar man uttrycka med att säga att å prioritätheten i (1) (som brukar benämnas beta-tätheten) är *konjugerad* till binomialmodellen eller Barnoullimodellen. Betatätheten är alltså konjugerad till binomialmodellen.

Accepterar man denna Bayesianska metodik, bör man efter det att de n oberoende försöken gjorts, anse att

$$p \approx \frac{a + f}{a + b + n}$$

eftersom å posteriori-tätheten antar sitt största värde i denna punkt. Säkerheten i å priori-gissningen angavs ju av $a + b$. Analogt anger $a + b + n$ säkerheten i å posteriori-gissningen.

Man kan tänka sig att säkerheten i å priori-gissningen $p \approx a/(a + b)$ är lika stor som den säkerhet man får om man i $a + b$ st oberoende mätningar ser händelsen A inträffa i a st. Men observera att a, b ej behöver vara heltalsvärda. Det enda kravet är att de är strikt positiva, d.v.s, $a > 0, b > 0$.

Undertecknad föredrar att på ett pragmatisk sätt blanda Bayesianska och frekventistiska idèer. Således anger han gärna följande å priori-konfidensintervall och -skattning

$$p = \hat{p}_{ML} \pm t_{\alpha/2}(a + b - 1) \sqrt{\hat{p}_{ML}(1 - \hat{p}_{ML})/(a + b - 1)}$$

där

$$\hat{p}_{ML} = \frac{a}{a+b}$$

(ML står för "Maximum Likelihood" och vi ska tänka på $\hat{p}_{ML} = a/(a+b)$ som p :s troligaste värde). Frihetsgradstalet $a+b-1$ bör rundas av nedåt om man ska hämta det ur en tabell och det ej är ett heltal. Ett implicit krav här är då att $a+b$ ej är för litet. Om detta ej är uppfyllt så får han avstå från att specificera ett a priori-intervall för p . I analogi anger han sedan a posteriori-konfidensintervallet och -skattningen till

$$p = \hat{p}_{ML} \pm t_{\alpha/2}(a+b+n-1) \sqrt{\hat{p}_{ML}(1-\hat{p}_{ML})/(a+b+n-1)}$$

där nu

$$\hat{p}_{ML} = \frac{a+f}{a+b+n}$$

Även här ska frihetsgradstalet $a+b+n-1$ vid behov rundas av nedåt. Det kan finnas skäl för att arbeta med ensidiga konfidensintervall. Så den möjligheten ska inte uteslutas. Observera att detta sätt att räkna ut konfidensintervall bara duger om antalet försök n är tillräckligt stort. Då bör inte $a+b$ vara för litet om det ska vara någon mening med den Bayesianska ansatsen. Men oftast är det nog så att Bayesiansk teknik gör bäst nytta när man inte har möjlighet att ta stora stickprov.

Observera att den pragmatiska Bayesianska metodiken, så som jag beskrivit den här och använder den, rymmer den klassiska frekventistiska. Väljer man $a=b=0$, blir a priori-tätheten konstant: $\pi(p) \propto 1$; och konfidensintervall och ML-skattning ovan blir de vi är vana vid. En konstant "prior" kallas *icke-informativ* och svarar mot ingen förhandskunskap.

Traditionella Bayesianer föredrar dock att skatta p med väntevärdet m.a.p a posteriori-tätheten. Då måste den ospecificerade konstanten bestämmas. Sedan ska man beräkna

$$\hat{p}_B = E[p | n, f] = \int_0^1 p \pi(p | n, f) dp$$

Exempelvis Beta kan assistera i denna process. En stunds bläddrande och tänkande ger vid handen att *Bayes-skattningen* av p är

$$\hat{p}_B = \frac{a+f+1}{a+b+n+2}$$

Notera att $\hat{p}_B = E[p | n, f]$ är en funktion av data n, f . Man kan visa att av alla reella funktioner $\Theta(n, f)$, så är Bayesskattningen $\hat{p}_B = E[p | n, f]$ den som minimerar väntevärdet

$$E [(\Theta(n, f) - p)^2 | n] = \int_0^1 \sum_{f=0}^n (\Theta(n, f) - p)^2 \pi(f | n, p) \pi(p) dp$$

Notera att $\pi(f | n, p) \pi(p) = \pi(f, p | n)$ är en två-dimensionell täthet och att väntevärdet ovan är m.a.p denna.

Traditionella Bayesianer föredrar s.k *kreditibilitetsintervall* framför konfidensintervall. Man väljer då intervallgränser l_1, l_2 så att

$$P(l_1 \leq p \leq l_2 | n, f) = \int_{l_1}^{l_2} \pi(p | n, f) dp = 1 - \alpha \quad (3)$$

Önskas ett uppåt begränsat intervall väljes $l_1 = 0$. Önskas ett symmetriskt intervall, så väljes l_1, l_2 så att

$$\int_0^{l_1} \pi(p | n, f) dp = \int_{l_2}^1 \pi(p | n, f) dp = \alpha/2$$

Då är ju också (3) uppfyllt.

Notera att det är den betingade fördelningen för f , givet n, p , som är $\text{Bin}(n, p)$. Fördelningen för f har tätheten

$$\pi(f | n) = \int_0^1 \pi(f | n, p) \pi(p) dp$$

ty produkten $\pi(f | n, p) \pi(p)$ är (i analogi med vanlig betingning) lika med den gemensamma fördelningen för paret f, p :

$$\pi(f | n, p) \pi(p) = \pi(f, p | n)$$

Tätheten $\pi(f | n)$ brukar kallas för den *prediktiva tätheten* för f . Det kan ibland finnas skäl att studera denna innan de n oberoende försöken utförs.

Antag nu att vi gjort våra n mätningar, uppdaterat vår à priori-kunskap $\pi(p)$ till à posteriorin $\pi(p | n, f)$ och önskar göra ytterligare m oberoende försök. Låt g vara antalet lyckade (d.v.s antalet gånger som A inträffar). Det är nu naturligt att som à prioritäthet välja $\pi(p | n, f)$. Vi kan nu räkna ut den *prediktiva tätheten* för g , givet n, f , enligt

$$\pi(g | m, n, f) = \int_0^1 \pi(g | m, p) \pi(p | n, f) dp$$

ty

$$\pi(g | m, p) \pi(p | n, f) = \pi(g, p | m, n, f)$$

där $\pi(g | m, p)$ är $\text{Bin}(m, p)$ -tätheten (mätmodellen) och $\pi(p | n, f)$ alltså vår nya à prioritäthet.

Efter det att de m nya försöken har gjorts och vi observerat frekvensen g kan vi räkna ut vår nya à posterioritäthet om vi som à prioritäthet väljer $\pi(p | n, f)$. Vi får icke oväntat

$$\begin{aligned} \pi(p | m, g, n, f) &\propto \pi(g | m, p) \pi(p | n, f) \\ &\propto p^{a+f+g} (1-p)^{b+n+m-(f+g)} \end{aligned}$$

Läsaren ombedes att själv tänka igenom följande "case:" Antag att man avser att testa i 10 timmar vardera 5 st exemplar av en vidareutveckling av en ventil som används i Formula-1-motorer. Syftet är att skatta funktionssannolikheten p efter 10 driftstimmar. Man är tämligen säker på att den ventil som används för närvarande har funktions-sannolikheten $p \approx 0.93$ och att vidareutvecklingen är en förbättring. Så man tänker sig à priori att $p \approx 0.93$. Väljer man att jobba Bayesianskt med en $\text{beta}(a, b)$ -à prioritäthet, så följer att $a/b \approx 0.93/0.07 \approx 13.3$. För att komma vidare behöver man någon slags skattning av säkerheten i antagandet $p \approx 0.93$. Ett sätt är att ansätta att $P(p \geq 0.9) \approx 0.8$. Lite "lek" med Matlab (jag använde funktionen `betacdf`) ger nu att $a = 40, b = 3$ kan vara ett vettigt val av betaparametrarna, ty då är à priori

$$\hat{p}_{ML} = \frac{40}{43} \approx 0.930 \quad \text{och} \quad P(p \geq 0.9) \approx 0.805$$

Icke oväntat visade det sig att alla komponenterna höll. À posteriori gäller då att

$$\hat{p}_{ML} = \frac{40 + 5}{40 + 3 + 5} = \frac{45}{48} = 0.938 \quad \text{och} \quad P(p > 0.9 | n = 5, f = 5) = 0.862$$

Men, hade t.ex endast en av komponenterna hållit, hade man istället fått ut

$$\hat{p}_{ML} = \frac{40 + 1}{40 + 3 + 5} = \frac{41}{48} = 0.854 \quad \text{och} \quad P(p > 0.9 | n = 5, f = 1) = 0.186$$

av à posteriori-analysen. Fast, med ett så dåligt utfall av försöket hade man nog miss-tänkt att något blivit helt fel i prototypen.

Sens moral av "caset" är att det är inte så lätt att påvisa förbättring när man redan har en hög funktionssannolikhet.

Exempel 11.1 A nonrepairable valve is assumed to have constant failure rate λ ...

Man avser att testa en komponent i sänder och göra en Bayesiansk uppdatering av kunskapen om λ efter varje.

Vi börjar med bestämma oss för att

$$\pi(\lambda) \propto \lambda^a e^{-\lambda s}$$

definierar en lämplig à priori-täthet för λ . Orsaken är att om man i ett stickprov bestående av a enheter observerar sammanlagda funktionstiden s tidsenheter, så blir troligheten precis $l(\lambda) = \lambda^a e^{-\lambda s}$. Ur t.ex formelsamlingen ser vi att $\Gamma(\alpha, \beta)$ -tätheten är

$$f(\lambda) = \frac{1}{\Gamma(\alpha) \beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \quad \text{för} \quad \lambda > 0$$

Så à priori är $\lambda \sim \Gamma(\alpha, \beta)$, med $\alpha = a + 1$ och $\beta = 1/s$. Att Gammafördelningen är konjugerad till exponentialmodellen ska vi strax se.

Sedan ska vi fixera parametrarna a, s . Høyland & Rausand väljer att sätta $\alpha = 2, \beta = 1$, vilket svarar mot $a = 1, s = 1$, alltså att man tidigare gjort $a = 1$ observation och erhållit resultatet $s = 1$. ML-skattningen av λ är därför à priori

$$\hat{\lambda}_{ML} = \frac{a}{s} = 1$$

Notera också att motsv Bayes-skattning är

$$E[\lambda] = \alpha\beta = \frac{a+1}{s} = 2$$

Sedan väntar vi på den första observationen t_1 och när vi har den gör vi vår första à posteriori-analys. Vi får att à posteriori-tätheten ges av

$$\pi(\lambda | t_1) \propto \lambda^{a+1} e^{-\lambda(s+t_1)}$$

M.a.o $\lambda \sim \Gamma(a+2, 1/(s+t_1))$. Låt oss anta att $t_1 = 1.3$. À posteriori är då ML-skattningen

$$\hat{\lambda}_{ML} = \frac{a+1}{s+t_1} = \frac{2}{2.3} = 0.87$$

och Bayes-skattningen

$$E[\lambda | t_1] = \frac{a+2}{s+t_1} = \frac{3}{2.3} = 1.30$$

Enl formelsamlingen är

$$2\lambda(s+t_1) \sim \chi^2(2(a+1))$$

om vi i $a+1$ försök har observerat totala testtiden $s+t_1$. Detta kan användas till ett konfidensintervall för λ om vi tänker frekventistiskt. Vill man istället ha ett kredibilitetsintervall för λ använder man att à posteriori är

$$\lambda | t_1 \sim \Gamma(a+2, 1/(s+t_1)) \Leftrightarrow 2\lambda(s+t_1) | t_1 \sim \chi^2(2(a+2))$$

Sedan använder vi à posteriori-tätheten, som à priori-täthet i analysen vi gör när vi har två observationer. Anta att $t_2 = 0.9$. Då är

$$\hat{\lambda}_{ML} = \frac{a+2}{s+t_1+t_2} = \frac{3}{3.2}$$

och

$$E[\lambda | t_1, t_2] = \frac{a+3}{s+t_1+t_2} = \frac{4}{3.2} = 1.25$$

Etc. □

Exempel 11.3 Anta att vi observerar en Poissonprocess med intensitet λ och har sett att $N(t) = n$. Då är modellen

$$\pi(n | \lambda, t) = P(N(t) = n | \lambda) = e^{-\lambda t} (\lambda t)^n / n!$$

Vi ser att även här är gammatätheten en lämplig a priori-täthet. Alltså,

$$\pi(\lambda) \propto \lambda^a e^{-\lambda s}$$

Å posterioritätheten ges då av

$$\pi(\lambda | t, n) \propto \lambda^{a+n} e^{-\lambda(s+t)}$$

Etc. □

Exempel 11.1 (Forts) Antag nu att vi satte igång n komponenter och har väntat tills vi sett r st fallera. Då har vi typ-II-censurerade data. För troligheten gäller

$$l(\lambda) \propto \lambda^r e^{-\lambda(t_{(1)} + \dots + t_{(r)})} e^{-\lambda(n-r)t_{(r)}} = \lambda^r e^{-\lambda(\sum_{i=1}^r t_{(i)} + (n-r)t_{(r)})} = \lambda^r e^{-\lambda \mathcal{T}_r}$$

Vi ser att även i denna situation är Gamma-fördelningen en lämplig a priorifördelning. Vi väljer alltså

$$\pi(\lambda) \propto \lambda^a e^{-\lambda s}$$

och för mätmodellen gäller enl ovan att

$$\pi(t_{(1)}, \dots, t_{(r)} | n, r) \propto \lambda^r e^{-\lambda \mathcal{T}_r}$$

Således gäller att

$$\pi(\lambda | n, r, t_{(1)}, \dots, t_{(r)}) \propto \lambda^{a+r} e^{-\lambda(s + \mathcal{T}_r)}$$

Etc. □