

Tommy Norberg
Matematisk statistik
Chalmers & GU
8 mars 2007

Riskkostnader, kostnads-nyttokalkyler och värdet av information

1 Introduktion

I vetenskapliga såväl som i vardagliga sammanhang används ordet "risk" på flera sätt. Ordet förknippas t.ex ofta med chansen (sannolikheten) att något obehagligt inträffar. Häromåret orsakade ett forskningsteam stor uppståndelse, då de menade att risken en fotgängare tar då han/hon ska korsa en gata, är förhöjd vid markerade övergångsställen. Förklaringen skulle kunna vara att fotgängare vid övergångsställen har en större benägenhet att slarva med uppsikten över bil- och cykeltrafiken. Fast man tycker ju att detta mer än väl borde kompenseras av ökad uppmärksamhet hos bilförare vid övergångsställen. Medicinare talar ofta om förhöjd risk för individer med ett visst beteende. Rökare har t.ex högre risk att få vissa typer av cancersjukdomar. Feta har större risk än normalviktiga att få åldersdiabetes. Etc.

I tekniska sammanhang har ordet risk mer och mer kommit att förknippas med en förväntad eller skattad framtida kostnad. Skälet till detta är bl.a att om man ska jämföra olika typer av möjliga olyckor, så är det inte bara deras resp sannolikheter som är av intresse utan också deras konsekvenser. De senare försöker man då uppskatta kostnaden för. Detta är, som var och en förstår, ingen lätt uppgift.

Det är viktigt att vänja sig vid att ordet risk har flera betydelser. Alla, inkl undertecknad, använder ordet risk i flera betydelser. Det är förståeligt att öva sig i att förstå ordet rätt.

Man måste också komma ihåg att de siffror som ingår i riskkalkyler ofta är mer eller mindre grova uppskattningar, alternativt uträknade eller framsimulerade i någon matematisk eller matematisk statistisk modell med (ibland stor) osäkerhet. En osäkerhetsanalys behöver alltså bifogas många riskanalyser. Detta görs inte alltid. Vi ger definitivt inte osäkerheten i analyserna tillräcklig uppmärksamhet i detta häfte.

Vi ska i våra tekniska värderingar av ordet risk också se till konsekvenskostnaden. Vi ska börja med den enklaste tolkningen av risk.

2 Riskkostnad

Man brukar definiera *risken* eller, snarare, *riskkostnaden* R som $C_F P_F$, där P_F är sannolikheten för "failure", som man ofta säger, F och C_F är den kostnad man får om F inträffar. Alltså,

$$R = C_F P_F \quad (1)$$

Här följer ett par övningar.

1. I en miljökonsekvens-analys av en existerande industrianläggning, har man bedömt att risken att en viss tungmetall-förorening når en närliggande vattentäkt till ca 1 på 1 000. Man vet att om tungmetall påträffas i täkten, så kommer folkopinionen att medföra att den inte utan åtgärd kan användas, även om halten tungmetall är under gränsvärdet. Låt oss anta att man då måste övergå till en annan vattentäkt. Kostnaden för detta är naturligtvis stor, säg att den är 1 500 Mkr.¹

(a) Bestäm risken.

Antag att det går att säkra vattentäkten till en kostnad av ca 1 Mkr. Helt säkert är ju inget, men man bedömer att sannolikheten att tungmetall når fram till täkten då minskar i alla fall med en faktor 10.

(b) Beräkna risken om täkten säkras. Hur stor är riskminskningen?

Vattentäkten kan förstöras/skadas också av andra orsaker. Anta t.ex att risken att en annan förorening (med mindre allvarliga konsekvenser) når fram (eller snarare påvisas) i vattentäkten är ca 1 på 50. Kostnaden för att åtgärda denna förorening skattas till 2.5 Mkr.

(c) Bestäm risken som denna harmlösare förorening ger upphov till.

(d) Hur stor är den totala risken täkten är exponerad för?

2. Ett järnvägsspår ska anläggas utmed en älv. Man bedömer att den årliga sannolikheten för en översvämning så kraftig att tågtrafik omöjliggörs är ca 1/10. Kostnaden, om detta sker, för det trafikerande bolaget beräknas till ca 250 kkr. Hur stor är den årliga risken?

¹1 Mkr = 1 000 000 kr, 1 kkr = 1 000 kr

3 Kostnad kontra nytta

Man kan göra en enkel s.k *kostnads-nyttokalkyl* i samband med situationen i övning 1, (a) och (b) ovan. Man jämför då kostnader och nyttor för olika handlingsalternativ. I situationen ovan är riskminskningen som handlingen ger upphov till lika med dess nytta. Vi har i (handlings-) alternativ 0 (gör inget) kostnaden 0 och nyttan (riskminskningen) 0. I alternativ 1 (säkra tåkten) har vi kostnaden 1 Mkr och riskminskningen 1.35 Mkr (j.f.r facit). Genomförs detta handlingsalternativ skulle man kunna tala om en ”vinst” för samhället om ca 0.35 Mkr. Det finns en teori för beslutsfattande som säger att det är rationellt (bäst) att genomföra alternativ 1, d.v.s att säkra tåkten. Så hade inte varit fallet om det hade kostat mer än 1.35 Mkr att säkra tåkten. Då hade alternativ 0 (gör inget) varit bäst.

Varje kostnads-nyttokalkyl innehåller ett noll-alternativ i vilket kostnader och nyttor är 0. Låt C_i och B_i vara kostnaden, resp nyttan i handlingsalternativ i . Man brukar definiera en s.k *objektfunktion*, som i föreliggande fall är

$$\Phi_i = B_i - C_i$$

för alternativ i . Här gäller att $B_i = R_0 - R_i$, så

$$\Phi_i = (R_0 - R_i) - C_i \quad (2)$$

I vårt enkla exempel är

$$\Phi_i = \begin{cases} 0 & \text{då } i = 0 \\ 0.35 & \text{då } i = 1 \end{cases}$$

ty $C_0 = 0$ och $C_1 = 1$ medan $R_0 = 1.5$ och $R_1 = 0.15$ (j.f.r facit).

Värdet av det bästa (eller de bästa, för det kan ju tänkas att mer än ett alternativ är bäst) handlingsalternativet är

$$\Phi = \max_i \Phi_i \quad (3)$$

Riskdefinitionen (1) gäller för engångsrisker. Det är tveksamt om den får användas i fall som det i övning 2. En i det sammanhanget bättre definition, som vi kommer att använda oss utav, fås om man tänker sig att failure inträffar enl en Poissonprocess

med intensitet λ_F st/år.² Den årliga ekonomiska risken (riskkostnaden) är då

$$R = C_F \lambda_F \quad (4)$$

(R är alltså den förväntade kostnaden på grund av failure under ett år. Obs att man naturligtvis kan räkna med andra tidsenheter än år.) I siffror blir slutresultatet densamma, eftersom sannolikheten för minst ett failure i Poissonprocessen är

$$P_F = 1 - e^{-\lambda_F} \approx \lambda_F$$

och approximationen är god då P_F är litet. Men observera att det kan ju faktiskt inträffa fler än ett failure under ett år även om sannolikheten för det är liten. Så det är helt klart vettigare att definiera återkommande risker på detta processbaserade sätt.

Gör nu följande övning.

3. Tänk dig en lastprocess som den i häftet om Poissonprocessen [2]. Antag att vi vet att tillbud T sker med intensiteten $\lambda_T \approx 0.15$, samt att styrkan av dessa är oberoende och fördelade enl Paretofördelning³

$$P(X > x|T) = 1 - F(x|T) = \left(\frac{x_T}{x}\right)^\alpha, \quad x \geq x_T$$

Vi ska anta att tillbudsnivån $x_T = 5.0$ och att $\alpha \approx 7.1$. Antag dessutom att katastrofnivån är $x_K = 7.0$ och vidare att katastroferna i lastprocessen kostar 100 kkr/st.

- (a) Beräkna risken.
- (b) Genom att vid varje tillbud aktivt bearbeta lastprocessen räknar man med att kunna reducera $P(K|T)$ till hälften. Denna aktivitet kostar 5 kkr/tillbud. Vad blir nu risken?
- (c) Gör en kostnads-nyttokalkyl för att se om det är lönsamt att aktivt bearbeta lastprocessen enl ovan.
- (d) Hur mycket får det kosta att reducera $P(K|T)$ till hälften, om det ska vara lönsamt?
- (e) Genom att vid tillbud koppla in ytterligare en maskin kan man höja katastrofnivån till 10. Vad blir i så fall intensiteten i katastrofsprocessen?

²När medicinare talar om risk är det ofta intensiteten i en Poissonprocess som avses

³Obs att denna definitionen av Paretofördelningen ej nödvändigtvis är den vi använde oss av i [2]. Det går därför inte att använda formlerna som härleddes i detta häfte rakt av

- (f) Antag att kostnaden för maskinen i ovanstående uppgift är 3.5 kkr per timma och att katastrof fortfarande kostar 100 kkr. Beräkna risken om tillbuden i snitt pågår i 2.9 timmar. (Det är tillåtet att räkna som om varje tillbud kostar $2.9 \cdot 3.5 = 10.15$ kkr.)
- (g) Är maskinen lönsam? Gör en kostnads-nyttokalkyl.

I den näst sista del-uppgiften i övning 3 ovan är failurekostnaden stokastisk. Man kan visa att även i sådana fall kan man räkna som om den vore fix. Det gäller alltså att uttrycket

$$R = C_F \lambda_F \quad (5)$$

för riskkostnaden är korrekt även om failurekostnaden är stokastisk. Det som är viktigt är att förekomsten av failure följer en Poissonprocess och att failurekostnaderna är oberoende av när i tiden de inträffar (d.v.s av Poissonprocessen) och att kostnader hörande till olika failure är oberoende stokastiska variabler med väntevärde C_F .

Bevis (svårt): Låt $X_1, X_2, \dots$ vara de resp failurekostnaderna. Antag att de är oberoende av varandra och av Poissonprocessen samt att $E[X_i] = C_F$ för $i = 1, 2, \dots$. Totala failurekostnaden under en tidsenhet blir

$$X = \sum_{i=1}^N X_i = \begin{cases} 0 & \text{då } N = 0 \\ X_1 & \text{då } N = 1 \\ X_1 + X_2 & \text{då } N = 2 \\ \vdots & \end{cases} = \sum_{n=0}^{\infty} \left(\sum_{i=1}^n X_i \right) 1_{\{N=n\}}$$

eller

$$X = \sum_{n=0}^{\infty} X 1_{\{N=n\}} = \sum_{n=0}^{\infty} \left(\sum_{i=1}^n X_i \right) 1_{\{N=n\}}$$

där N , som representerar antalet failure under 1 tidsenhet, är Poissonfördelad med parameter λ_F . Här och i det följande: 1_A betecknar en stokastisk variabel som antar värdet 1 eller 0 beroende på om A eller A' inträffar. Den förväntade kostnaden på

grund av failure blir

$$\begin{aligned}
E[X] &= E \left[\sum_{n=0}^{\infty} \left(\sum_{i=1}^n X_i \right) 1_{\{N=n\}} \right] \\
&= \sum_{n=0}^{\infty} E \left[\left(\sum_{i=1}^n X_i \right) 1_{\{N=n\}} \right] \\
&= \sum_{n=0}^{\infty} E \left[\sum_{i=1}^n X_i \right] P(N = n) \\
&= \sum_{n=0}^{\infty} \sum_{i=1}^n E[X_i] P(N = n) \\
&= \sum_{n=0}^{\infty} n C_F P(N = n) \\
&= C_F \lambda_F
\end{aligned}$$

Här har använts i tur och ordning linjäritet, oberoende och linjäritet, samt att

$$\sum_{n=0}^{\infty} n C_F P(N = n) = E[C_F N] = C_F \lambda_F \quad \text{Q.E.D}$$

I samhällsekonomiska beslutssituationer landar man ofta i kostnads-nyttokalkyler där nyttor och kostnader sträcker sig över flera år. Det kan då vara på sin plats att diskontera. Då är

$$\Phi_i = \sum_{t=0}^T \frac{\Phi_i(t)}{(1+r)^t} \quad (6)$$

där $\Phi_i(t) = B_i(t) - C_i(t)$ är värdet av objektfunktionen för alternativ i t år från nu och r är den s.k. *diskonteringsräntan*, som här antas vara konstant i tiden, men som inte behöver vara det. Man kan kalla Φ_i för *nuvärdet av alternativ i* och $\Phi = \max_i \Phi_i$ för *nuvärdet*. Poängen med att diskontera blir tydlig i fall då man har kostnader för ett projekt långt innan nyttorna kommer.

Inspiration till nästa övning har författaren fått från [4]. Innan kan det vara idé

att påminna om formeln för summan av en ändlig geometrisk serie:

$$\sum_{t=0}^T \rho^t = \frac{1 - \rho^{T+1}}{1 - \rho}$$

4. Ett land funderar på att med början nästa år vidta klimatrelaterade åtgärder, vars kostnad uppgår till ca 10 000 Mkr/år under 20 år. Man tror att dessa åtgärder kommer att reducera skadeeffekter om ca 8 000 Mkr/år under 75 år med början 21 år från nu. Beräkna nuvärdet om diskonteringsräntan sätts till
- (a) 0
 - (b) 2%
 - (c) 4%

För vilka av de tre diskonteringsräntorna ovan är det lönsamt för samhället att vidta de klimatrelaterade åtgärderna?

Man kan tänka sig negativa diskonteringsräntor.⁴ Man kan även tänka sig att ha med två diskonteringsräntor i kalkylerna. En för nyttorna och en för kostnaderna. En viktig orsak till att diskontera på kostnadssidan är att samhället blir rikare — 10 miljarder är en betydligt större belastning av samhällsekonomin idag jämfört med om 20 år. I alla fall om samhället fortsätter att successivt bli rikare (vilket många tror och räknar med ska ske). Det är tveksamt om samma argument ska användas på nytto-sidan om dessa består utav minskningar av risk eller skadeverkningar på miljön. I situationen som betraktades i övning 4 ovan kan man tänka sig en diskonteringsränta för kostnaderna om ca 2–4% och en lägre räntenivå för nyttorna. Den senare kanske t.o.m ska väljas aningen negativ.

Vid överbelastning upphör många konstruktioner att fungera. Tänk dig att överbelastning inträffar enl en Poissonprocess med intensitet λ_F st/tidsenhet, att de tidsrymder konstruktionen är ur drift är oberoende stokastiska variabler med väntevärde μ och att kostnaden för ett stopp är proportionellt mot stoppets längd. Om proportionalitetskonstanten är C kkr/tidsenhet, så blir riskkostnaden enl vad som sagts ovan $R = \lambda_F C \mu$ kkr/tidsenhet.

⁴Ätminstone kan författaren till detta häfte tänka sig det

Obs att detta är en approximation som förutsätter att $\mu \ll 1/\lambda_F$, d.v.s $\lambda_F\mu \ll 1$. Konstruktionen är ju inte under risk då den är ur drift. Den verkliga failure-intensiteten är

$$\omega_F = \frac{1}{\frac{1}{\lambda_F} + \mu} = \frac{\lambda_F}{1 + \lambda_F\mu}$$

och riskkostnaden är, enl resonemanget i föregående stycke,

$$R = \omega_F C \mu = \frac{\lambda_F C \mu}{1 + \lambda_F\mu}$$

Gör nu följande övning.

5. I en produktionsanläggning har man ca 1 stopp per månad p.g.a komponentbrist. Man avser att göra en kostnads-nyttokalkyl för att se om det är kostnads-effektivt att reducera frekvensen sådana stopp. Vi börjar med att titta på nu-läget. Det är rimligt att räkna med ca 20 arbetsdagar per månad och varje arbetsdag består av 8 timmars effektivt arbete. Om man får ett stopp p.g.a komponentbrist, så kan inte arbetet återupptas förrän vid nästa arbetsdags start. Man räknar med att arbets- och maskinkostnaden samt kostnaden för utebliven försäljning uppgår till ca 3 000 kr/timme.

- (a) Hur mycket kostar dessa stopp per månad?

Man kan, genom att tillämpa en annan beställningsstrategi, reducera antalet stopp av denna art till ca 1 per 1.5 månader.

- (b) Hur mycket får detta lov att kosta? Gör en kostnads-nyttanalys för att se var gränsen för lönsamhet går.

Många gånger har man (åtminstone i teorin) hyfsad koll på sannolikheten för failure, dvs att det går fel. En inte ovanlig situation är att man medelst simulering kan uppskatta denna sannolikhet. Viktigt att hålla i minnet är då att resultatet bara är giltigt om simuleringsmodellen beskriver verkligheten korrekt.

Gör nu nästa övning.

6. Antag att man i en serie av 10 000 simuleringar erhöll failure F i 73 av dem. Skatta sannolikheten för failure, P_F , och beräkna ett uppat begränsat konfidensintervall för densamma. Konfidensgraden ska vara ca 95%.

I många situationer har man en à priorisannolikhet för failure F , som uppdateras efter undersökning(ar). Här följer en sådan övning. Låt D vara händelsen att undersökningen resulterar i att failure F är trolig. Då D inträffar säger vi att failure har *detekterats* i undersökningen (som ju kan vara en regelrätt mätning).

7. Ett icke helt ovanligt (orimligt) sätt att modellera total okunskap är att ansätta att sannolikheten för failure är $P(F) = 0.5$. Antag nu att man gör vissa undersökningar. Erfarenhetsmässigt vet man att den betingade sannolikheten att detektera failure givet failure är $P(D|F) = 0.8$ samt att feldetekteringssannolikheten är $P(D|F') = 0.1$.
 - (a) Bestäm sannolikheterna för failure givet detektion resp ingen detektion.
 - (b) Antag nu att man inte detekterade failure och gör ytterligare ett försök att detektera failure med samma detekterings- och feldetekterings-sannolikheter. Bestäm sannolikheterna för failure givet detektion resp ingen detektion. (Tänk på att nu är det inte längre rimligt att ansätta $P(F) = 0.5$.)

Anta att det kostar C_R att "investera bort" 80% av risken $C_F P(F)$. Bästa alternativet är då värt

$$\Phi = \max(0, 0.8 C_F P(F) - C_R)$$

och vi ser att $\Phi > 0$ då, och bara då,

$$0.8 C_F P(F) > C_R \quad \Leftrightarrow \quad P(F) > 1.25 C_R / C_F \quad (7)$$

Små risker har man alltså (med detta ekonomiska synsätt) ingen anledning att åtgärda.

Fundera ordentligt över svaret till den avslutande frågan i nästa övning.

8. För en potentiellt förorenad tomt gäller att man bedömer att samhällets kostnad om tomten är förorenad och den ej saneras är ca 4 gånger saneringskostnaden, som vi för enkelhets skull antar är ca 1 Mkr. Efter diskussion med experter görs bedömningen att risken att tomten verkligen är förorenad är liten, ca 1 på 5 eller 0.2. Om detta stämmer är det ej lönsamt att sanera tomten.
 - (a) Verifiera detta påstående.

Men beslutsfattaren känner sig osäker på den låga skattningen av failure-sannolikheten, så hon bestämmer sig för att göra en mätning. Hon har att välja på två tillvägagångssätt—en billig fältnätning, i vilken riskerna för fel är stora och en dyrare, i vilken ett antal prover sänds till ett ackrediterat laboratorium för analys och som har lägre felsannolikheter. Oavsett metod, väljer man att testa statistiskt så att risken att felaktigt klassa tomten som ren ej är större än ca 0.05. För den billiga metoden gäller att sannolikheten för en falsk detektering är stor, ca 0.75. För den dyrare metoden gäller att motsv sannolikhet är mindre, ca 0.5. Vi ska nu undersöka vad som händer i de bägge fallen om förorening detekteras (j.f.r med terminologin i övning 7).

- (b) Antag att det var den billiga metoden som användes. Är det nu lönsamt att remediera? Hur stort är Φ ?
- (c) Antag att det var den dyrare metoden som användes. Är det nu lönsamt att remediera? Hur stort är Φ ?

Besvara i båda fallen följande fråga. Var det någon mening med att mäta?

Då

$$\Phi = \max(0, 0.8 C_F P(F) - C_R)$$

såg vi att remediering är lönsamt om $P(F) > 1.25 C_R/C_F$, så $P(F) = 1.25 C_R/C_F$ är den sannolikhet för vilken beslutsosäkerheten är maximal. När $P(F)$ är ungefär så här stor, finns det ett behov av att skaffa sig ytterligare information. I det som följer ska vi gå igenom en metodik för att värdera nyttan av ny information.

4 Värdet av ny information

I det som följer är det praktiskt att tänka på Φ som à priori- eller prior-värdet av bästa alternativet, och vi betecknar det Φ_{prior} . D.v.s,

$$\Phi_{\text{prior}} = \max(0, C_F P(F) - C_R) \quad (8)$$

(Här och i det som nu följer, tänker vi oss att man till kostnaden C_R helt kan ta bort risken $C_F P(F)$.) Antag nu att vi, som i övningarna 7 och 8 ovan, gör vissa undersökningar i syfte att bättre skatta sannolikheten för failure. Låt D beteckna händelsen att failure F har detekterats. Vi antar att sannolikheterna för fel av första resp andra slaget är kända. De betecknas $P(D|F')$ och $P(D'|F)$. Vi

definierar à posteriori- eller posterior-värdet

$$\Phi_{\text{post}} = \begin{cases} \max(0, C_F P(F|D') - C_R) & \text{om } D' \text{ inträffar} \\ \max(0, C_F P(F|D) - C_R) & \text{om } D \text{ inträffar} \end{cases} \quad (9)$$

Innan vi genomför undersökningarna kan vi såklart inte veta utfallet av dem och således inte heller posteriorvärdet, men vi kan beräkna förväntat posteriorvärde, ofta kallat preposteriorvärdet,

$$\begin{aligned} \Phi_{\text{prepost}} &= E[\Phi_{\text{post}}] \\ &= \max(0, C_F P(F|D') - C_R)P(D') + \max(0, C_F P(F|D) - C_R)P(D) \end{aligned} \quad (10)$$

Differensen

$$W = \Phi_{\text{prepost}} - \Phi_{\text{prior}} \quad (11)$$

är (preposterior-) *värdet* av informationen eller data (som man ofta säger). Det är naturligt att förutsätta att sambandet mellan F och D är positivt.⁵ Notera dock att under alla omständigheter måste $W \geq 0$.

Bevis: Självklart gäller att preposteriorvärdet är icke-negativt, d.v.s

$$\Phi_{\text{prepost}} \geq 0$$

Vi har dessutom att preposteriorvärdet är minst

$$\begin{aligned} \Phi_{\text{prepost}} &\geq (C_F P(F|D') - C_R)P(D') + (C_F P(F|D) - C_R)P(D) \\ &= C_F P(F \cap D') - C_R P(D') + C_F P(F \cap D) - C_R P(D) \\ &= C_F P(F) - C_R \end{aligned}$$

Härur följer

$$\Phi_{\text{prepost}} \geq \max(0, C_F P(F) - C_R) = \Phi_{\text{prior}} \quad \text{Q.E.D}$$

Man kan visa att $W > 0$ om, och endast om, data har potential att ändra bästa beslutet (j.f.r övning 8 ovan).

Gör nu följande övning.

⁵Händelserna F och D sägs vara *positivt korrelerade* om $P(F|D) \geq P(F)$

9. I sanering av förorenad mark är det vanligt att dela in i lämpligt stora s.k. remedieringsenheter, som saneras vid behov och en i sänder. I en tänkt tillämpning (j.f.r [3]) gäller att underlåtenhet att sanera en förorenad enhet kostar ca 4 gånger saneringskostnaden, som är ca 1 Mkr. Man tror dessutom att $P(F) = 0.75$. Antag att $P(D|F) = 0.8$ och att $P(D|F') = 0.15$.

- (a) Hur stort är priorvärdet?
- (b) Hur stort är preposteriorvärdet?
- (c) Hur stort är datavärdet W ?
- (d) Om istället $P(F) = 0.5$, vad blir då W ?
- (e) Om istället $P(F) = 0.25$, vad blir då W ?

Ibland behöver man ta till mer avancerade Bayesianska metoder för att uppskatta (eller uppdatera) sannolikheten för failure. J.f.r [1].

10. Tio mätningar av en viss sjös surhet ska göras i avsikt att avgöra om den behöver kalkas eller ej och man vet från andra sjöar i omgivningen att sannolikheten att pH-värdet understiger det aktuella gränsvärdet är $p_c \approx 0.3$. Man bedömer att denna förhandsinformation är ungefär lika mycket värd som informationen från de 10 mätningarna som ska göras.

- (a) Föreslå en lämplig a priorifördelning för p_c .
- (b) Regeln är att sjön ska kalkas om minst 6 av de 10 mätningarna understiger gränsvärdet. Bestäm a priorisannolikheten för kalkning.
- (c) Det visade sig att gränsvärdet understegs i 5 av de 10 mätningarna. Bestäm a posterioriskattningen av sannolikheten p_c .

Antag att kostnaden för att kalka sjön i övning 10 ovan är C_R . Ett för lågt pH-värde påverkar sjöns ekologi negativt. Eko-ekonomer uppskattar kostnaden för detta till C_F . De uppskattar även att kalkning reducerar denna kostnad med en faktor 10. En naturlig kostnads-nyttokalkyl utmynnar i

$$\begin{aligned}\Phi_{\text{prior}} &= \max \left(0, \left(C_F P(F) - \frac{1}{10} C_F P(F) \right) - C_R \right) \\ &= \max \left(0, \frac{9}{10} C_F P(F) - C_R \right)\end{aligned}$$

och vi ser att kalkning är lönsamt bara då $P(F) > \frac{10}{9} C_R / C_F$.

Gör nu nästa övning.

11. (J.f.r övning 10 ovan.) Låt $P(F) = E[p_c]$, där vi antar att $p_c \sim \text{beta}(4, 8)$. Antag dessutom att $C_F = 10C_R$. Bestäm Φ_{prior} . Är kalkning lönsamt? Om istället $C_F = 2C_R$, är kalkning då lönsamt?

Vi fortsätter nu med att beräkna värdet av det i övning 10 ovan föreslagna mätprogrammet. Låt X vara antalet observationer under gränsvärdet. Notera att

$$X|p_c \sim \text{bin}(n, p_c) \quad \text{där} \quad n = 10, \quad p_c \sim \text{beta}(4, 8)$$

Härur följer att $E[X|p_c] = np_c$ samt att $E[X] = nE[p_c] = n/3$. Notera vidare att

$$P(F|x) = E[p_c|x] = \frac{4+x}{12+n}$$

Således gäller att

$$\begin{aligned} \Phi_{\text{post}} &= \max \left(0, \frac{9}{10} C_F P(F|x) - C_R \right) \\ &= \max \left(0, \frac{9}{10} \frac{4+x}{12+n} C_F - C_R \right) \end{aligned}$$

Obs att $\Phi_{\text{prior}} = \Phi_{\text{post}}|n=0$. Antag, som i övning 11 ovan, att $C_F = 10C_R$. Då

$$\Phi_{\text{post}} = \max \left(0, \left(\frac{9(4+x)}{12+n} - 1 \right) C_R \right) \quad \text{och} \quad \Phi_{\text{prior}} = 2C_R$$

Det är lätt att se att $\frac{9(4+x)}{12+n} > 1$ för alla x om $n = 10$. Då gäller följaktligen att

$$\Phi_{\text{post}} = \left(\frac{9(4+x)}{22} - 1 \right) C_R \quad \Rightarrow \quad \Phi_{\text{prepost}} = 2C_R$$

och vi ser att $W = 0$. Det föreslagna mätprogrammet är alltså värdelöst.

Orsaken till detta är att risken uttryckt i C_F är stor. Låt oss anta att den är mindre, säg att $C_F = 2C_R$. Då gäller att

$$\Phi_{\text{prior}} = 0$$

medan

$$\Phi_{\text{post}} = \max \left(0, \left(\frac{9(4+x)}{5(12+n)} - 1 \right) C_R \right)$$

Obs att även här gäller att $\Phi_{\text{prior}} = \Phi_{\text{post}}|n = 0$. Insättning av $n = 10$ ger

$$\Phi_{\text{post}} = \max \left(0, \left(\frac{9(4+x)}{110} - 1 \right) C_R \right)$$

Det är lätt att se att $\Phi_{\text{post}} > 0$ bara då $x \geq 9$. Härur inses att $W > 0$, samt att kalkning bör ske bara då minst 9 av de 10 mätningarna ligger under gränsvärdet.

Ingen hänsyn togs i analysen ovan till att även det föreslagna mätprogrammet har en kostnad. Man bör kräva att datavärdet är minst lika stort som mätkostnaden för att mätning ska anses lönsam. Men obs att regeln som talar om när kalkning bör ske inte påverkas av mätkostnaden.

Facit

1. (a) 1.5 Mkr (b) 0.15 Mkr, 1.35 Mkr (c) 0.05 Mkr = 50 kkr (d) 1.55 Mkr
2. 25 kkr
3. (a) 1.376 kkr/dygn (b) 0.688 kkr/dygn (c) $\Phi_1 = 0.688 - 5 \cdot 0.15 = -0.062$ kkr/dygn eller -0.414 kkr/tillbud, det är alltså ej lönsamt (d) 4.586 kkr/tillbud (e) 0.001093 st/dygn (f) 1.632 kkr/dygn (g) nej, ty $\Phi_2 = 1.2665 - 0.15 \cdot 10.15 = -0.256$ kkr/dygn eller -1.706 kkr/tillbud
4. (a) 400 000 Mkr (b) 44 714 Mkr (c) $-49\,444$ Mkr
5. (a) 11 707 kr/månad (b) $\leq 3\,838$ kr/månad
6. $\hat{P}_F = 0.0073$, $P_F < 0.0087$
7. (a) $P(F|D) = 8/9 \approx 0.889$, $P(F|D') = 2/11 \approx 0.182$ (b) $P(F|D) = 16/25 = 0.64$, $P(F|D') = 4/85 \approx 0.047$
8. (a) $P(F) = 1/5 \not\geq 1/4 = C_R/C_F$ (b) nej ty $P(F|D) = 0.241 < 0.25$, $\Phi = 0$ (c) ja ty $P(F|D) = 0.322 > 0.25$, $\Phi = 0.288 > 0$
9. (a) 2 Mkr (b) 2 Mkr (c) 0 (d) 0.125 Mkr (e) 0.4875 Mkr
10. (a) $\text{beta}(4, 8)$ (b) ≈ 0.047 (c) $\hat{p}_c = 0.4$
11. $2C_R$, nej

Referenser

- [1] Tommy Norberg: Introduktion till Bayesiansk uppdatering. Häfte 2006.
- [2] Tommy Norberg: Poissonprocessen och extrema laster. Häfte 2005.
- [3] Tommy Norberg, Lars Rosén: Calculating the optimal number of contaminant samples by means of data worth analysis. *Environmetrics* 2006; **17**: 705–719.
- [4] Tore Söderqvist: Diskontering av samhällsekonomiska analyser av klimatåtgärder. Naturvårdsverket. Rapport 5618, 2006.