

HVORDAN TENKTE THOMAS BAYES?

Bent Natvig

Matematisk institutt, Universitetet i Oslo

La meg først takke for innbydelsen til å innlede ved dette emnekurset “Eit allmenne-medisinsk partitur” som en del av Nidaroskongressen '97. Det er helt storveis for en klart over gjennomsnittet hypokonder å ha så mye dyktig medisinsk personell som tilhørere og diskusjonspartnere.

Den engelske presten og sannsynlighetsteoretikeren Thomas Bayes ble født i 1702 og døde i 1761. Jeg skal ikke her gå inn i historiske kildematerialer for å belyse hvordan Bayes tenkte der og da. Snarere skal jeg forsøke å overbevise dere om at Bayesianisk statistikk er et viktig metodisk grunnlag for medisinske fag her og nå.

Mitt eget fagområde er pålitelighetsteori – som er en matematisk grunnlagsteori for risikoanalyser av store teknologiske systemer som oljerigger og kjernekraftverk. Typisk for slike systemer er at de er bygget opp av en lang rekke komponenter. Disse kan være teknologiske og menneskelige, men funksjonerer og feiler avhengig av hverandre. Pga. manglende data vil en ved faktiske risikoanalyser av teknologiske systemer være helt avhengig av å kunne utnytte skjønn og erfaringer fra ingeniører om de teknologiske komponenter, og fra psykologer og sosiologer om de menneskelige komponenter. Dette leder til at en må basere seg på en statistisk metodelære som også naturlig utnytter subjektive sannsynligheter. En slik metodelære er nettopp Bayesianisk statistikk.

Det er klare fellestrekk mellom pålitelighets- og risikoanalyse av teknologiske systemer og allmenne-medisin. Allmenne-medisinerens system er enkeltmennesket. Eksempler på komponenter her er organer og lemmer. Pålitelighetsanalytikerens og allmenne-medisinerens spesialfelt er helheten – ikke de enkelte komponenter. Avhengigheten mellom komponentene er spesielt viktig for å vurdere helheten. For teknologiske systemer skaper ytre stress som vær og vind slik avhengighet, mens for enkeltmennesket skapes slik avhengighet av psykisk stress i tillegg til det fysiske stress.

Videre er pålitelighetsanalytikerens og allmenne-medisinerens fokus ikke rettet mot å teste hypoteser om egenskaper til populasjoner av systemer/mennesker, men å treffe fornuftige beslutninger under usikkerhet om det enkelte system/menneske på grunnlag av best mulige riskovurderinger. I bunn for slike beslutninger bør det ligge en relevant teori. Det vil bli argumentert for at Bayesianisk beslutningsteori her er det helt naturlige fundament.

1 Hvordan tenker allmennmedisineren?

For at jeg skulle få et bedre inntrykk av hva som rører seg i miljøet, fikk jeg oversendt 6 utgaver av “UTPOSTEN – blad for allmenn- og samfunnsmedisin”. La meg som en outsider med en gang si at jeg ble meget imponert over bredden og kvaliteten på artiklene i dette bladet. Selv om jeg ikke fant noen diskusjon eller referanse til Bayesiansk statistikk i de artiklene jeg leste, følte jeg ofte at jeg var blant åndsfrender. Jeg skal gi noen eksempler på det.

La meg starte med noen sitater fra Petter Øgars utmerkede artikkel (1) om skjønnsutøvelse i samfunnsmedisinen:

“I en skjønnsutøvelse vil flere vurderinger inngå:

Hva er relevante faktorer og forhold å ta hensyn til ved den beslutning som skal treffes eller vurdering som skal gjøres?

Hvilken helsefaglig kunnskap finnes om disse faktorer og forhold, og hvilken vitenskapelig kvalitet er det på denne faglige dokumentasjonen?

Hvor sannsynlig er det at ulike faktorer og forhold skal inntreffe?

Hvilken verdi skal tillegges de ulike forhold?

En samlet bearbeiding og integrering av disse delvurderingene.

Hvem er kompetent til å foreta disse ulike vurderingene?”

“Samfunnsmedisinen er i sin natur forebyggende og handler om å påvirke morgendagens og de neste 5–50 års helsetilstand. Dette framtidsperspektivet må nødvendigvis innebære usikkerhetsmomenter. Kunnskapsgrunnlaget vårt bygger og må bygge på gårsdagens empiri og risikofaktorer. Vi trekker slutninger om dagens og morgendagens forhold på grunnlag av gårsdagens kunnskap og eksponeringer. Dagens miljøfarer som påvirker folks framtidshelse, kan vi ikke ha sikker kunnskap om. Kravet om faglighet i naturvitenskapelig forstand vil ofte være umulig eller meningsløst å oppfylle. Deler av samfunnsmedisinen er også av natur slik at man ikke på forhånd kan forutsi med samme grad av sannsynlighet som innenfor naturvitenskapene. Den vitale samfunnsmedisin må derfor være forut for sin tid og våge den usikkerhet og kontrovers som følger av føre-var-prinsippet. Det sentrale og ytterst skjønnspregede spørsmål er hvor føre var en skal tillate seg å være på hvilket grunnlag, og i denne vurderingen identifisere og vedkjenne seg de usikkerhetsmomenter som ligger i beslutningsgrunnlaget.

Vurderingen av kvaliteten på våre faglige kunnskaper er en viktig samfunnsmedisinsk ferdighet og er i seg selv en skjønnsutøvelse. Det er et betydelig behov for å erkjenne faglige utilstrekkeligheter og usikkerheter og legge dem åpent fram i dagen. Ikke som et uttrykk for faglig fiasko, men som uttrykk for faglig ærlighet og edruelighet. Vi bør i større grad synliggjøre hvor kunnskapen

slutter og de skjønsmessige vurderinger overtar. I mange sammenhenger vil samfunnsmedisinen være tjent med å legge fra seg en påtatt objektivisme.

I forhold til framtidrettede vurderinger er ikke skjønnsutøvelsen bare knyttet til kvalitetsvurdering av dagens kunnskaper. Den er også knyttet til sannsynlighetsvurderinger. Hvor stor sjanse er det for at en begivenhet inntreffer gitt at de faglige forutsetningene x og y gjelder? I beredskapsarbeid er det dette en prøver å gjøre gjennom såkalte risiko- og sårbarhetsanalyser.”

La meg hoppe til en annen interessant artikkel (2) av Linn Getz om den allmennmedisinske vitenskapen:

“Det er en gylden tid for såkalt Evidence-Based Medicine. Men hvis vi skal lite på de fakta som vi er så dristige at vi omtaler som bevis, må vi sikre oss at de framkommer i vår egen setting og har gyldighet for den.”

2 En anvendelse av Bayes teorem på oppdatering av sannsynligheter på grunnlag av ny informasjon

La meg starte med å sitere en siste meget relevant artikkel (3) av Steinar Hunskår og Eivind Meland som tar opp myten om det ufeilbarlige helsevesen.

“Ved undersøkelse av pasienter med mistenkt sykdom vil det alltid være usikkerhet: De fleste vet at sykdom kan foreligge selv med normale undersøkelser og testresultater (prøven var “normal”, men pasienten viste seg å være syk). Mindre kjent er det kanskje at undersøkelse og test-resultater kan være positive uten at pasienten er syk (prøven viser “sykdom”, men pasienten er frisk!). I det siste tilfellet snakker vi om “falsk positive”. Den medisinske usikkerhet og sjanse for feil er et uttrykk for et allment fenomen: vitenskapen kan ikke bevise noe, den kan bare sannsynliggjøre standpunktet med en viss fare for feil.

Når legen stiller en diagnose er det med andre ord ikke et uttrykk for en absolutt sannhet, det foreligger bare en viss sannsynlighet for at sykdommen foreligger. Omvendt vil diagnosen “frisk” også bare være sannsynliggjort med varierende grad av sikkerhet. Heldigvis er sannsynligheten for å ha rett ofte stor, spesielt ved viktige og alvorlige sykdommer der det er gjort mange forskjellige undersøkelser. Men vi må kunne vedstå oss at ved flere tilstander vil selv den beste lege stille korrekt diagnose i mindre enn 50% av tilfellene! Hvis vi ikke åpent tør vedstå oss dette, er vi offer for et bedrag som kan føre til at medisinen utvikles til en allmenn trussel i vår kultur – helsetjenester blir bjørnetjenester som truer folkehelsen.”

Det sannsynlighetsteoretiske verktøy som passer som hånd i hanske til disse problemstillingene, er Bayes teorem. Dette er det alminnelig enighet om uansett synspunkt på Bayesianisk statistikk. La oss illustrere dette ved følgende konstruerte eksempel. Anta at vi tester en pasient utfra mistanke om at vedkommende har en blodsykdom. Innfør følgende notasjon:

$$\begin{aligned} S &= \{\text{Pasienten er alvorlig syk}\} \\ L &= \{\text{Pasienten er litt syk}\} \\ N &= \{\text{Pasienten er helt frisk}\} \\ + &= \{\text{Testen gir positivt utslag}\} \\ - &= \{\text{Testen gir negativt utslag}\} \end{aligned}$$

Anta at vi utfra generell norsk helsestatistikk vet at 2% av befolkningen er alvorlig syke, 10% er litt syke mens altså da 88% er helt friske. La $P(S)$ betegne sannsynligheten for at pasienten er alvorlig syk med tilsvarende definisjoner for $P(L)$ og $P(N)$. Utfra helsestatistikken over kan vi anta, hvis vi ikke utnytter mer subjektiv tilleggsinformasjon om pasienten, at

$$P(S) = 0.02 \quad P(L) = 0.10 \quad P(N) = 0.88 .$$

Anta på basis av et større datamateriale at vi har følgende kunnskap om testens egenskaper. I 90% av tilfellene gir testen korrekt positivt utslag når den anvendes på en alvorlig syk pasient. Tilsvarende gir testen positivt utslag i 60% av tilfellene når den anvendes på en litt syk pasient, mens den helt feilaktig gir positivt utslag i 10% av tilfellene når den anvendes på en helt frisk pasient. La $P(+|S)$ betegne den betingete sannsynlighet for at testen gir positivt utslag gitt at pasienten er alvorlig syk med tilsvarende definisjoner for $P(+|L)$ og $P(+|N)$. Utfra informasjonen over kan vi anta at:

$$P(+|S) = 0.90 \quad P(+|L) = 0.60 \quad P(+|N) = 0.10 .$$

Vi ønsker nå å beregne sannsynligheten for at pasienten vår virkelig er alvorlig syk gitt at testen har gitt positivt utslag, dvs. vi er på jakt etter $P(S|+)$. Det er her Bayes teorem fra sannsynlighetsregningen er som skreddersydd. La $P(S \cap +)$ betegne sannsynligheten for S "snitt" $+$, dvs. sannsynligheten for at både pasienten er alvorlig syk og at testingen av ham gir positivt utslag med tilsvarende definisjoner for $P(L \cap +)$ og $P(N \cap +)$. Utfra definisjonen på betinget sannsynlighet har vi at:

$$P(S|+) = \frac{P(S \cap +)}{P(+)} ,$$

der $P(+)$ er sannsynligheten for at testen gir positivt utslag på en uspesifisert pasient. Det burde være lett å overbevise seg om at:

$$P(+)= P(S \cap +) + P(L \cap +) + P(N \cap +) ,$$

siden en uspesifisert pasient kan tilhøre en av følgende gjensidig utelukkende grupperinger: de alvorlig syke, de litt syke eller de helt friske. Ved å bruke definisjonen på betinget sannsynlighet en gang til har vi at:

$$\begin{aligned}P(S \cap +) &= P(+|S)P(S) \\P(L \cap +) &= P(+|L)P(L) \\P(N \cap +) &= P(+|N)P(N) .\end{aligned}$$

Samler vi dette, får vi at:

$$P(S|+) = \frac{P(+|S)P(S)}{P(+|S)P(S) + P(+|L)P(L) + P(+|N)P(N)} .$$

Dette er nettopp Bayes teorem anvendt på vår problemstilling. Ved å sette inn de numeriske verdier for sannsynlighetene på høyre side av likhetstegnet får vi at:

$$P(S|+) = \frac{0.90 \cdot 0.02}{0.90 \cdot 0.02 + 0.60 \cdot 0.10 + 0.10 \cdot 0.88} = 0.11 .$$

Det er følgelig en sannsynlighet så lav som 0.11 for at pasienten vår virkelig er alvorlig syk gitt at testen har gitt positivt utslag. Dette er selvsagt betenkelig lavt noe allmennmedisineren må ta hensyn til i vurderingen av og kommunikasjonen med pasienten. Enkelte vil hevde at testen rett og slett er ubrukelig.

En kan imidlertid bøte på situasjonen ved å foreta en ytterligere, uavhengig test for å skaffe mer informasjon. Dette svarer til A og B prøver i dopingtesting. Begge disse må være positive for å gi domfellelse. Vi ønsker nå å beregne sannsynligheten for at pasienten vår virkelig er alvorlig syk gitt at begge testene har gitt positivt utslag. La oss innføre følgende notasjon:

$$\begin{aligned}+_1 &= \{\text{Første test gir positivt utslag}\} \\+_2 &= \{\text{Andre test gir positivt utslag}\} .\end{aligned}$$

Vi er følgelig på jakt etter $P(S|+_1 \cap +_2)$.

Betrakt situasjonen etter at første test er utført og har gitt positivt utslag. I stedet for å ta utgangspunkt i $P(S)$, $P(L)$ og $P(N)$, basert på generell norsk helsestatistikk, starter vi nå med de oppdaterte sannsynligheter $P(S|+_1)$, $P(L|+_1)$ og $P(N|+_1)$ basert på informasjonen at første test av vår pasient ga positivt utslag. Ved Bayes teorem beregnet vi over at $P(S|+_1) = P(S|+) = 0.11$. Tilsvarende finner en at:

$$\begin{aligned}P(L|+_1) &= \frac{0.60 \cdot 0.10}{0.90 \cdot 0.02 + 0.60 \cdot 0.10 + 0.10 \cdot 0.88} = 0.36 \\P(N|+_1) &= \frac{0.10 \cdot 0.88}{0.90 \cdot 0.02 + 0.60 \cdot 0.10 + 0.10 \cdot 0.88} = 0.53 .\end{aligned}$$

Uavhengigheten av testene kan fortolkes på følgende måte. Hvis vi skal beregne sannsynligheten for at andre test gir positivt utslag gitt at pasienten er alvorlig syk,

avhenger ikke denne av resultatet av den første testen. Dermed har vi at:

$$\begin{aligned}P(+_2|S \cap +_1) &= P(+_2|S) = P(+|S) = 0.90 \\P(+_2|L \cap +_1) &= P(+_2|L) = P(+|L) = 0.60 \\P(+_2|N \cap +_1) &= P(+_2|N) = P(+|N) = 0.10 ,\end{aligned}$$

som er de samme sannsynlighetene som inngikk i beregningen av $P(S|+)$ over.

Dermed erstatter vi bare $P(S)$, $P(L)$ og $P(N)$ med $P(S|+_1)$, $P(L|+_1)$ og $P(N|+_1)$ og får igjen ved å bruke Bayes teorem at:

$$P(S|+_1 \cap +_2) = \frac{0.90 \cdot 0.11}{0.90 \cdot 0.11 + 0.60 \cdot 0.36 + 0.10 \cdot 0.53} = 0.27 .$$

Dette er vesentlig bedre enn 0.11, men fortsatt altfor lavt.

Hovedpoenget med disse beregningene er imidlertid å påvise hvor perfekt Bayes teorem passer til å oppdatere sannsynligheter når ny informasjon dukker opp. Dette er helt ukontroversielt i beregningene her fordi all informasjon er basert på “håndfaste” data. Bakoversveisen hos enkelte statistikere oppstår hvis utgangspunktet er subjektiv tilleggsinformasjon om pasienten i stedet for generell norsk helsestatistikk. Igjen kan en stille spørsmålsteget ved “påattatt objektivisme”.

3 En anvendelse av Bayesiansk beslutningsteori

La oss forfølge vår pasient et skritt videre etter at begge testene har gitt positivt utslag. Allmennmedisineren må nå fatte en beslutning på grunnlag av de oppdaterte sannsynligheter. Vi har at $P(S|+_1 \cap +_2) = 0.27$. Tilsvarende finner en at:

$$\begin{aligned}P(L|+_1 \cap +_2) &= \frac{0.60 \cdot 0.36}{0.90 \cdot 0.11 + 0.60 \cdot 0.36 + 0.10 \cdot 0.53} = 0.59 \\P(N|+_1 \cap +_2) &= \frac{0.10 \cdot 0.53}{0.90 \cdot 0.11 + 0.60 \cdot 0.36 + 0.10 \cdot 0.53} = 0.14 .\end{aligned}$$

På grunnlag av de to testene er det klart mest sannsynlig at pasienten bare er litt syk, hvilket ikke skulle tilsi nærmere og mer kostbar utredning i denne omgang. På den annen side er det en ikke ubetydelig sannsynlighet for at pasienten faktisk er alvorlig syk og dermed omgående bør innlegges på sykehus. Allmennmedisineren står overfor en beslutning under stor usikkerhet. Skal pasienten innlegges eller ei? Innfør følgende notasjon svarende til de to alternativene:

$$\begin{aligned}B_1 &= \{\text{pasienten hospitaliseres omgående}\} \\B_2 &= \{\text{en ser det hele an inntil videre}\} .\end{aligned}$$

For å fatte en beslutning kommer en ikke utenom å *subjektivt* anslå tap av nytte ved å velge B_1 eller B_2 i forhold til pasientens faktiske tilstand S , L eller N . Hvis en

f.eks. hospitaliserer en helt frisk pasient, representerer dette bortkastede ressurser for samfunnet, forlenget ventetid i sykehuskø for andre pasienter, tapt arbeidsinnsats for pasientens eventuelle arbeidsgiver og endelig i beste fall bortkastet tid og ressurser for pasienten. La meg igjen sitere den sistnevnte artikkel (3):

“Generelt kan personer som har fått påvist funn som ikke er uttrykk for sykdom, invalidiseres av angst og sosial forventning om sykerollen. Forskning har lært oss at den subjektive sykdomsopplevelse for slike pasienter kan være like sterk som blant dem med reell sykdom. Mange har fått uføretrygd på grunnlag av slik ikke-sykdom. Vi ser slike pasienter, men de er ikke egnet for sensasjonsoppslag i pressen.”

Hvis en på den annen side ser det hele an inntil videre, og pasienten er alvorlig syk, kan dette medføre invalidisering eller tidlig død med tilhørende omkostninger for familie og samfunn. Anta allmennmedisineren, som til syvende og sist skal fatte beslutningen, kommer frem til følgende tabell over tapt nytte målt i norske kroner.

		Pasientens tilstand		
		S	L	N
Beslutning	B_1	0	25000	50000
	B_2	100000	25000	0

Tabell over tapt nytte ved de to alternative beslutninger.

Vi ser av tabellen at hvis pasienten er alvorlig syk og allmennmedisineren helt korrekt fatter beslutningen om omgående hospitalisering, er tapt nytte lik 0. Dersom allmennmedisineren i denne situasjonen velger å se det hele an, gjøres den alvorligste feilen med tapt nytte kr. 100 000. Hvis pasienten er litt syk, vurderes de to alternative beslutningene å ha det samme nyttetapet på kr. 25 000,-. For tilfellet at pasienten er helt frisk og allmennmedisineren helt korrekt fatter beslutningen om å se det hele an, er igjen tapt nytte lik 0. Dersom allmennmedisineren i denne situasjonen velger å hospitalisere pasienten omgående, gjøres den nest mest alvorlige feilen med tapt nytte kr. 50 000.

Den vanlige matematiske definisjon av risiko ved en beslutning er forventet tapt nytte. Dette betyr at en må veie sammen tapt nytte ved de ulike tenkelige tilstander som kan oppstå, der vektene er de best mulig estimerte sannsynligheter for disse tilstandene.

Risikoen ved beslutningen om å hospitalisere pasienten omgående er følgelig:

Risiko ved B_1 =

$$\begin{aligned}
 &= 0 \cdot P(S|+_1 \cap +_2) + 25000 \cdot P(L|+_1 \cap +_2) + 50000 \cdot P(N|+_1 \cap +_2) \\
 &= 0 \cdot 0.27 + 25000 \cdot 0.59 + 50000 \cdot 0.14 = 21750
 \end{aligned}$$

Tilsvarende er risikoen ved å se det hele an inntil videre:

$$\text{Risiko ved } B_2 = 100000 \cdot 0.27 + 25000 \cdot 0.59 + 0 \cdot 0.14 = 41750 .$$

Følgelig ser vi at forventet tapt nytte eller risikoen ved å se det hele an inntil videre, er nær dobbelt så stor som ved at pasienten omgående hospitaliseres. Dermed er det opplagt at allmennmedisineren bør treffe den siste beslutningen selv om den best mulige estimerte sannsynlighet for at pasienten er alvorlig syk er så lav som 0.27. Det er også klart at vi kan gjøre relativt store endringer i tabellen over tapt nytte ved de to ulike beslutninger, og likevel ende opp med en beslutning om at pasienten hospitaliseres omgående.

4 Bayesiansk beslutningsteori generelt

Bayesiansk statistikk og beslutningsteori, samt andre former for beslutningsteori, er behandlet i (4). Vi skal belyse litt av denne Bayesianske teorien i det følgende. Nøkkelparameteren vi er interessert i, betegnes ofte med den greske bokstaven θ . Den kan gjerne være en vektor av flere størrelser. Nøkkelparameteren er fast, men ukjent for beslutningstageren. Beslutningstageren formulerer sin usikkerhet om parameteren før data er samlet inn ved hjelp av en såkalt *à priori sannsynlighetstetthet* (*punktsannsynlighet*) betegnet med $\pi(\theta)$.

I gjennomgangseksemplet vårt er nøkkelparameteren, θ , pasientens helsetilstand. Den kan anta tre verdier θ_1, θ_2 og θ_3 gitt ved:

$$\theta_1 = S, \quad \theta_2 = L, \quad \theta_3 = N.$$

Den tilhørende *à priori* punktsannsynlighet er gitt ved:

$$\begin{aligned}\pi(\theta_1) &= P(S) = 0.02 \\ \pi(\theta_2) &= P(L) = 0.10 \\ \pi(\theta_3) &= P(N) = 0.88.\end{aligned}$$

En samler så inn data betegnet med D . Sannsynlighetstettheten (punktsannsynligheten) for disse dataene, *gitt nøkkelparameteren* θ , betegnes $f(D|\theta)$. Som funksjon av θ betegnes denne med $L(\theta)$, dvs.

$$L(\theta) = f(D|\theta).$$

$L(\theta)$ kalles *rimelighetsfunksjonen* (engelsk likelihood function). I gjennomgangseksemplet vårt har vi at de første data er at testen gir positivt utslag, dvs.:

$$D = +.$$

Rimelighetsfunksjonen er der gitt ved:

$$\begin{aligned}L(\theta_1) &= f(D|\theta_1) = P(+|S) = 0.90 \\ L(\theta_2) &= f(D|\theta_2) = P(+|L) = 0.60 \\ L(\theta_3) &= f(D|\theta_3) = P(+|N) = 0.10.\end{aligned}$$

Vi er nå på jakt etter usikkerheten i nøkkelparameteren etter at dataene D er samlet inn. Denne usikkerheten er gitt ved en såkalt *à posteriori sannsynlighetstetthet* (*punktsannsynlighet*) betegnet med $\pi(\theta|D)$. Denne finnes ved Bayes teorem, og er rett og slett

proporsjonal med produktet av rimelighetsfunksjonen, $L(\theta)$, og à priori sannsynlighetstettheten (punktsannsynligheten), $\pi(\theta)$. Følgelig har vi:

$$\pi(\theta|D) = L(\theta)\pi(\theta)/K .$$

K er her en konstant som sikrer at $\pi(\theta|D)$ virkelig er en sannsynlighetstetthet (punktsannsynlighet). Følgelig har vi hvis θ kan anta verdier i $\Theta = \{\theta_1, \theta_2, \dots\}$ at:

$$K = L(\theta_1)\pi(\theta_1) + L(\theta_2)\pi(\theta_2) + \dots$$

Hvis θ kan anta verdier i en mengde, Θ , av reelle tall, må vi erstatte summasjon med integrasjon og får at:

$$K = \int_{\Theta} L(\theta)\pi(\theta)d\theta .$$

I vårt gjennomgangseksempel får vi f.eks. at:

$$\pi(\theta_1|D) = \frac{L(\theta_1)\pi(\theta_1)}{L(\theta_1)\pi(\theta_1) + L(\theta_2)\pi(\theta_2) + L(\theta_3)\pi(\theta_3)} = 0.11 .$$

Får en nye data, gjentar en prosedyren over men med ny à priori sannsynlighetstetthet (punktsannsynlighet) gitt ved den gamle à posteriori sannsynlighetstettheten (punktsannsynligheten). Det var nettopp det vi gjorde i gjennomgangseksemplet der de nye data er at også den andre testen gir positivt utslag.

Anta vi nå skal velge mellom k mulige beslutninger B_1, B_2, \dots, B_k . Beslutningstageren må da subjektivt anslå tap av nytte ved valg av beslutningene B_1, \dots, B_k i forhold til den sanne verdi av nøkkelparameteren θ . Dette tapet av nytte er gitt ved *tapsfunksjonen* $L(\theta, B_i)$ som gir tapet av nytte ved å fatte beslutningen B_i når nøkkelparameteren antar verdien θ . Tapsfunksjonen er følgelig en funksjon av to variable. I gjennomgangseksemplet vårt fremstilte vi denne ved hjelp av en tabell.

Risikoen, eller *à posteriori forventet tapt nytte*, ved beslutningen B_i er gitt ved:

$$L(\theta_1, B_i)\pi(\theta_1|D) + L(\theta_2, B_i)\pi(\theta_2|D) + \dots ,$$

hvis θ kan anta verdier i $\Theta = \{\theta_1, \theta_2, \dots\}$. Hvis θ kan anta verdier i en mengde, Θ , av reelle tall, må vi erstatte summasjon med integrasjon og får i stedet

$$\int_{\Theta} L(\theta, B_i)\pi(\theta|D)d\theta .$$

Risikoen ved beslutning B_i er følgelig et veiet gjennomsnitt av tapt nytte for de ulike verdier av θ , der vektene er gitt ved à posteriori sannsynlighetstettheten (punktsannsynligheten) $\pi(\theta|D)$.

À posteriori Bayes beslutningen er da den beslutning blant $B_i, i = 1, \dots, k$ som har lavest risiko.

Generelt vil tapsfunksjonen alltid måtte vurderes subjektivt av beslutningstageren. På denne bakgrunn er det overraskende at mange stritter i mot å trekke inn subjektivt skjønn også i sannsynlighetsvurderingene som en slik risikoanalyse baserer seg på.

5 Bayesianske metoder i medisin

Jeg presenterte noen hovedidéer i Bayesiansk beslutningsteori på et seminar som Sosial- og helsedepartementet arrangerte 9. desember i fjor “Klinisk forskning – er de statistiske arbeidsmåter gode nok?”. La oss hente følgende utdrag fra innledningen til Statsråd Gudmund Hernes (5):

“Disse problemene ligger i grenselandet mellom metode og moral, statistikk og etikk. Noen av disse problemene fikk brå aktualitet ved den debatt vi hadde om høydosebehandling i sommer, men de spenner videre og stikker dypere.”

“For en som ser medisinen utenfra, er det slående å registrere i hvilken grad man bygger på den klassiske eller frekventistiske tilnærming til statistikk. Formålet med den klassiske statistikk er først og fremst å teste hypoteser. Eller sagt på en annen måte: å dikotomisere resultatene etter hvorvidt de er signifikante eller ikke.

Det som ellers er medisinsens kjerneproblem, er å treffe beslutninger, særlig beslutninger under usikkerhet. Her vil også klinikerne mene at man i tillegg til generell viten også må vektlegge klinisk skjønn. Skal man f.eks. gi høydosebehandling, kan man ikke bare bygge på estimater om hvor hyppig slik behandling lykkes – en rekke andre trekk ved pasientens tilstand må vurderes i tillegg.

Bayesiansk statistikk er utviklet for å ta hensyn til skjønn og (voksende) tilleggsinformasjon på en måte som er forenlig med sannsynlighetsregningens regler. Slik skulle den passe særlig godt til medisinske beslutninger som nettopp er beslutninger under usikkerhet.

Sagt på en annen måte: I klassisk statistikk undersøker man om to behandlinger A og B er like gode. Klassisk statistikk sier at vi skal velge behandling A hvis A er så mye bedre enn B at det er urimelig å tro at forskjellen skyldes tilfeldigheter. Jo mer urimelig det er å tro på tilfeldigheter, desto mer signifikant er forskjellen mellom behandling A og B.

Bayesiansk statistikk, derimot begynner med den observerte forskjellen mellom A og B og spør hvor stor sjanse det er for at A faktisk er bedre enn B gitt den observerte forskjellen.

Klassikerne spør altså hva sjansen er for å få den observerte forskjell ved rene tilfeldigheter hvis det *ikke* er noen forskjell.

Bayesianerne spør hva er sannsynligheten for at det skal være en underliggende reell forskjell gitt den forskjellen vi faktisk observerer. Man starter med det man vet eller tror, og undersøker hvor mye ny informasjon har forbedret kunnskapen.

Konklusjonene som trekkes ved disse to ulike tilnærmingene er ikke alltid den samme. Og likevel er det den klassiske statistikk som rår grunnen blant norske klinikere – ja, i stor grad også internasjonalt, selv om ekspertene i matematisk statistikk i økende grad er blitt Bayesianere.

Om jeg formulerer det spisst: Den klassiske kliniske forskning velger en statistisk hovedretning som reduserer effekten av den fornuft som ligger i klinikernes skjønn.

Eller hvis jeg sier det på en annen måte: I medisin er problemet mer å treffe beslut-

ninger enn å teste hypoteser. Da må hele tiden utgangspunktet være det en *vet* eller har gode grunner for å tro – selv om det er skjønnsmessig – *på forhånd*. Det et forsøk gir, er informasjon som setter en i stand til å *justere* det en trodde. Resultatene fra den ene undersøkelse danner grunnlaget for den neste.

Spørsmålet blir derfor: Hva er grunnen til at norske medisinerere i så liten grad bygger sine undersøkelser på en Bayesiansk tilnærming?”

Rett etter departementsseminaret fikk jeg sammen med Ivar Aursnes, Institutt for farmakoterapi, UiO et oppdrag fra departementet. Oppdraget var å gi en vurdering basert på Bayesiansk statistikk av den vitenskapelige dokumentasjon av legemidlene etidronat (Didronate) og alendronat (Fosamax) for reduksjon av antall brudd hos kvinner med osteoporose. Det ble utført også i samarbeid med mine kolleger Geir Storvik og Jørund Gåsemyr. Som evalueringskriterier benyttet vi først og fremst sammenligning av de relative effektivitetsforbedringer i forhold til placebo for de to legemidlene. Nøkkelparametrene for hvert enkelt studie som vi baserte oss på, var:

p_1 = Sannsynligheten for minst en fraktur i løpet av 3 år når placebo er brukt

p_2 = Sannsynligheten for minst en fraktur i løpet av 3 år når legemiddelet er brukt

og

λ_1 = Intensitet av frakturer per år når placebo er brukt

λ_2 = Intensitet av frakturer per år når legemiddelet er brukt.

Helt sentralt i den Bayesianske metodikk generelt er som vi har sett, å beskrive usikkerheten i slike nøkkelparametre ved hjelp av sannsynlighetsfordelinger. Denne usikkerhet blir beskrevet også etter at data er samlet inn, dvs. når beslutninger skal fattes, og ikke før som ved klassiske konfidensintervall.

Innsikten i parametrenes sanne verdier før data er tatt hensyn til, ble i dette oppdraget beskrevet ved såkalte ikke-informative à priori sannsynlighetsfordelinger. I praksis spilte det her liten rolle hvilken à priori innsikt som ble lagt til grunn da datagrunnlaget var svært omfattende. På bakgrunn av data ble disse fordelinger oppdatert til à posteriori sannsynlighetsfordelinger. Dette gjør det mulig å evaluere og ta hensyn til usikkerheten som ligger i effektivitetsforbedringene på en enklere og mer konsistent måte. Datamateriale fra ulike studier er koblet sammen for å gi en mer pålitelig vurdering av legemidlene. Denne sammenkoblingen er igjen gjort ved hjelp av Bayesiansk statistikk. Endelig ble det foretatt en kostnadsvurdering av de to legemidlene som igjen er basert på Bayesiansk statistikk. En kompliserende faktor ved studiene som var tilgjengelige, var at forekomsten (prevalensen) av symptomgivende osteoporose i populasjonene som er testet, varierer til dels svært mye. En kombinerende av studiene er likevel mulig å utføre hvis en gjør en antagelse om at den relative effektivitetsforbedring av legemiddelet er den samme for alle nivåer av forekomsten. Dette betyr at vi kan konsentrere oss om usikkerheten i de avledete nøkkelparametrene:

$$\alpha = (p_1 - p_2)/p_1, \quad \beta = (\lambda_1 - \lambda_2)/\lambda_1,$$

som er felles for de underliggende studiene.

British Medical Journal hadde 7. september i fjor en artikkel (6) "Bayesian statistical methods. A natural way to assess clinical evidence", skrevet av Laurence Freedman – Acting chief. Artikkelen er absolutt verdt å lese. Her henvises også til en annen lærebok (7). Dette inspirerte teamet nevnt over som utførte Departementets oppdrag, til å sende en engelsk oversettelse av den norske rapporten (8) til dette tidsskriftet. Denne fant de så interessant at vi ble bedt om å utarbeide en publiserbar artikkel på grunnlag av rapporten. Dette arbeider vi med nå. Uansett skjebne til akkurat dette arbeidet – Bayesianiske metoder er på full fart inn i statistikkfaget i sin alminnelighet og i medisinsk statistikk spesielt. Forsøkene i visse kretser på skrubbing kan bare skotøybransjen glede seg over.

Litteratur:

1. Øgar P. Skjønnsutøvelse i samfunnsmedisinen. Utposten 1996; 25:7–8.
2. Getz, L. Den allmenntilleggs vitenskapen. Utposten. 1995; 24:7.
3. Hunskår, S., Meland E. Myten om det ufeilbarlige helsevesen (1). Utposten 1995; 24:1.
4. Berger, J.O. Statistical decision theory and Bayesian analysis. Second edition. Springer-Verlag, 1985.
5. Hernes, G. Innledning ved Sosial- og helsedepartementets departementsseminar 9. september: Klinisk forskning – er de statistiske arbeidsmåter gode nok?, 1996.
6. Freedman, L. Bayesian statistical methods, A natural way to assess clinical evidence. British Medical Journal 1996; 313: 569–70.
7. Berry, D.A. Statistics. A Bayesian perspective. Duxbury Press, 1996.
8. Storvik, G., Natvig, B., Gåsemyr, J., Aursnes, I. Vurdering basert på Bayesianisk statistikk av den vitenskapelige dokumentasjon av legemidlene etidronat (Didronate) og alendronat (Fosamax) for reduksjon av antall brudd hos kvinner med osteoporose. Rapport utarbeidet for Sosial- og helsedepartementet, 1997.