

Opprinnelig publisert i boken *Naturvitere filosoferer* (Roger Strand, Glenn A. Bristow, redaktører), Megaloceros forlag, Bergen 1997, side 14–28.

Lagt ut på Internett av Ivar Heuch, Matematisk institutt, Universitetet i Bergen, i august 2008, under adressen <http://www.uib.no/People/nmaih/arkiv/kap/bayesfrek2008.pdf>

Striden mellom bayesianere og frekventister om idégrunnet for statistiske slutninger

Ivar Heuch

Vi statistikere blir ofte oppsøkt av brukere av faget som vil ha råd om valg av statistisk metode i praktiske situasjoner. I mange tilfeller blir brukerne forbauset når vi antyder at flere vidt forskjellige angrepsmåter kan være aktuelle, og at de ulike metodene kan tenkes å føre til avvikende konklusjoner. Av og til hender det at selve grunnlaget for de statistiske slutningene også blir berørt. Da må statistikerne innrømme at det ikke en gang foreligger noen allmenn enighet om de fundamentale prinsippene.

Ved statistisk analyse av datasett vil man vanligvis søke å ta hensyn til usikkerheten som ligger i de enkelte observasjonene. Uansett hvilke prinsipper som legges til grunn, blir denne usikkerheten bygget inn i en modell for de innsamlede dataene ved hjelp av begrep fra sannsynlighetsteorien. F. eks. vil en måling av en fysisk konstant som regel falle i nærheten av den sanne, underliggende verdien med høy sannsynlighet. Det kan brukes modeller som i større eller mindre grad spesifiserer størrelsen av slike sannsynligheter, avhengig av konkret problemstilling og mer fundamental holdning til prinsippene for statistiske slutninger. Det er likevel et fellestrekk ved modellene at de uttaler seg om sannsynligheter knyttet til de mulige observasjonsverdiene på en måte som viser at de aktuelle sannsynlighetene avhenger av de underliggende korrekte verdiene.

Hensikten med en statistisk analyse er normalt å uttale seg om slike ukjente størrelser som ligger bak observasjonene, men som vi likevel ikke kan si noe helt sikkert om på grunnlag av dataene. En standardbetegnelse for disse størrelsene i faget statistikk er *ukjente parametere*. Det kan dreie seg om fysiske konstanter som ikke kan måles helt nøyaktig, eller f. eks. andeler av den norske befolkning som har bestemte politiske sympatier, belyst ved data fra begrensede utvalgsundersøkelser. Statistiske metoder bør føre frem til utsagn om de ukjente parametrene, med større eller mindre grad av sikkerhet, på grunnlag av observasjonene. I mange problemstillinger kan en ønske å bestemme konfidensintervaller som parametrene ligger i med stor sikkerhet, eller det kan være aktuelt å teste hypoteser om parametrene med utgangspunkt i dataene. Utsagnene som blir formulert, må i alle fall bygge på sannsynlighetsmodellen for observasjonene, med varierende sannsynligheter avhengig av de sanne parameterverdiene.

Den store uenigheten oppstår når man skal prøve å formulere prinsippene for konstruksjon av statistiske metoder og samtidig beskrive usikkerheten som forplanter seg til utsagnene om de ukjente parametrene. Grovt sagt finnes det to konkurrerende retninger i striden om prinsippene for metodevalg. *Bayesianerne* vil søke å uttrykke *all* usikkerhet ved å stille opp utsagn som bygger på sannsynligheter, også ved vurdering av usikkerheten for bakenforliggende parametre. Når modellbeskrivelsen angir sannsynligheter for begivenheter knyttet til observasjonene, gitt de sanne parameterverdiene, regner bayesianerne seg frem til en slags “inverse sannsynligheter” for parametrene, gitt de bestemte verdiene som er observert. *Frekventistene* reserverer derimot bruken av sannsynlighetsbegrepet til beskrivelsen av usikkerheten i dataene selv. De utvikler statistiske metoder der graden av usikkerhet i konklusjonene blir målt på forskjellig måte, mer direkte ut fra sannsynlighetene gitt ved modellbeskrivelsen for observasjonsverdiene.

Striden mellom bayesianerne og frekventistene er på ingen måte av ny dato. Den bayesianske tankegang har sitt utspring i et arbeid av den engelske presten Thomas Bayes (1702–1761) publisert posthumt i 1763. Arbeidet viste i datidens matematiske språk bl. a. hvordan man i en bestemt situasjon kunne regne ut “inverse sannsynligheter.” Fremgangsmåten ble etter hvert populær i mer generelle sammenhenger, og ble spesielt utviklet som allment hjelpemiddel av Laplace. Som en regneteknisk metode utgjør det tilsvarende *Bayes’ teorem* om “betingede sannsynligheter” en grunnleggende setning i våre dagers elementærkurs i sannsynlighetsregning. Selve regneregelen er slett ikke omstridt. Det er derimot bruken av regelen for å fremkaste utsagn om ukjente parametre ved analyse av datasett.

Den frekventistiske tankegangen henger i stor grad sammen med fremveksten av statistisk inferensteori fra 1920-årene av, særlig i forbindelse med engelskmannen R. A. Fishers banebrytende arbeider. Med dagens utstrakte bruk av statistiske metoder på mange områder er den frekventistiske tankegangen dominerende. Nær sagt alt som læres bort i grunnleggende kurs i statistikk, enten det dreier seg om hypotesetesting eller konstruksjon av konfidensintervaller, bygger på frekventistiske idéer. Bayesianske metoder er imidlertid populære innenfor bestemte avgrensede anvendelsesområder, og i de siste ti årene har en kunnet iaktta en ny og mer åpen holdning til slike metoder, iallfall blant fagstatistikerne.

Få emner har i årenes løp gitt opphav til slik strid i statistikernes rekke som diskusjonen for og mot bayesianske idéer. De frekventistiske metodene er ikke i samme grad basert på ett enkelt grunnprinsipp som de bayesianske, og mange frekventister vil være villige til å vurdere forskjellige grunnleggende angrepsmåter i ulike praktiske situasjoner. Talsmenn for den bayesianske retningen understreker derimot at deres prinsipielle innstilling er ment å dekke *alle* statistiske problemstillinger. Bayesianerne blir derfor ofte stående som den ekstreme part ved mange konfrontasjoner, noe som kan virke uheldig ved en mer generell sammenligning. De tilbakeviser bestemt antydninger om at det hersker en nesten religiøs innstilling i den bayesianske leiren, og at de selv driver misjonsvirksomhet (Lindley, 1978). Også blant norske statistikere har vi hatt

til dels kraftig meningsutveksling om emnet, bl. a. med en diskusjon startet av Natvig (1989), som ledet til reaksjoner fra Sverdrup (1990), Norberg (1991) og flere andre.

Bayes' teorem

Siden Bayes' teorem er så fundamentalt for forståelsen av den bayesianske tankegangen, vil vi forsøke å gi et inntrykk av innholdet i setningen ved hjelp av et eksempel. I vanlige kurs i sannsynlighetsregning blir setningen formulert med symboler for betingede sannsynligheter. For å gi grunnlag for en mer omfattende diskusjon vil vi heller forklare resonnetet bak Bayes' teorem direkte ut fra eksempelet. Senere vil dette eksempelet også bli brukt til å belyse forskjellen mellom bayesiansk og frekventistisk tankegang.

Eksempelet stammer i prinsippet fra Fisher (1956), men det har også blitt trukket frem ellers når bayesianske prinsipper har vært diskutert. Bl. a. er det videreført av Edwards (1972). Slik det presenteres her, krever eksempelet bare grunnleggende kjennskap til utregning av sannsynligheter basert på genetiske resonnementer. Det må riktignok innrømmes at eksempelet er satt opp av en frekventist, så det gir muligens ikke helt uhildet dokumentasjon. Det bør også understrekes at eksempelet lett kan gi et feilaktig inntrykk av den generelle relevans som statistiske problemstillinger har. Problemet som beskrives er valgt nettopp fordi det gir en enkel illustrasjon av de vesentlige idéene.

La oss betrakte et gen hos mus som påvirker fargen av pelsen, og som forekommer med to mulige alleler B og b . De diploide individene har én av tre mulige genotyper BB , Bb og bb . Som vanlig i slike resonnementer, skiller vi ikke mellom rekkefølgen av symbolene, så kombinasjonen bB er regnet med i tilfellet Bb . Allelet B for svart farge er dominant, noe som medfører at både den homozygote genotypen BB og heterozygoten Bb er svarte. Mus med homozygot genotype bb er derimot brune. I det følgende ser vi bort fra eventuelle muligheter for mutasjon av allelene B og b .

Vi tenker oss nå at to individer som vi *vet* er heterozygote Bb blir krysset. (Dette kan vi f. eks. få i stand ved å hente avkom etter en tidligere krysning mellom to individer fra linjer som langt tilbake bare har hatt henholdsvis svart og brun farge.) Ut fra vanlige genetiske regler kjenner vi sannsynlighetene for genotypene blant avkommet fra krysningen $Bb \times Bb$. Ved å betrakte de forskjellige kombinasjonene av kjønnsceller finner vi at genotypen BB har sannsynligheten $1/4$, mens genotypen Bb har sannsynligheten $1/2$, og genotypen bb har sannsynligheten $1/4$. De to første avkomskategoriene blir altså svarte mus, mens den siste kategorien utgjør brune mus.

Eksempelet dreier seg nå hovedsakelig om genotypen til en tilfeldig utvalgt *svart* mus fra denne krysningen. Vi kaller henne for enkelhets skyld Minnie. Dette individet må nødvendigvis ha en av de to genotypene BB eller Bb . Sannsynlighetene for de to tilfellene må ha verdier i forholdet $(1/4) : (1/2)$ eller $1 : 2$, og siden summen av sannsynlighetene for alle mulighetene skal være lik 1, er altså sannsynligheten $1/3$ for at Minnie skal ha genotypen BB og $2/3$ for at hun skal ha genotypen Bb .

Vi forestiller oss at Minnie blir krysset med en brun hanmus (med genotype bb). Det viser seg at vi får 7 svarte avkomsindivider og ingen brune. Dette utgjør nå vårt datamateriale. Spørsmålet er hvordan denne observasjonen innvirker på sannsynlighetene for de to mulige genotypene BB og Bb for Minnie. Hvis Minnie i virkeligheten er homozygot BB , må nødvendigvis *alle* avkomsindividerne fra krysningen $BB \times bb$ bli heterozygote Bb . Det stemmer for så vidt med vår observasjon. Er Minnie derimot heterozygot Bb , får et avkomsindivid fra krysningen $Bb \times bb$ hver av genotypene Bb og bb med sannsynlighet $1/2$. Muligheten av at Minnie er heterozygot, står derfor slett ikke i strid med de 7 observerte svarte avkomsindividerne, selv om den registrerte verdien tydeligvis virker mindre rimelig i dette tilfellet.

La oss sette opp sannsynlighetene for det vi faktisk har observert, under de to forskjellige mulighetene for Minnies genotype. Hvis Minnie har genotype Bb , blir sannsynligheten $(1/2)^7 = 1/128$ når vi forutsetter at genotypene hos avkomsindividerne bestemmes uavhengig av hverandre. Hvis Minnie derimot har genotypen BB , er sannsynligheten lik 1. På forhånd visste vi at de to genotypetilfellene for Minnie hadde sannsynlighetene $2/3$ og $1/3$. Tenker vi på sannsynligheter som relative frekvenser eller brøkdeler i det lange løp, ser vi at det er naturlig å multiplisere sammen de to verdiene $2/3$ og $1/128$ for å finne sannsynligheten for at Minnie er heterozygot Bb og at hun samtidig skal bli mor til 7 svarte mus. Dette tilfellet har derfor sannsynligheten $(2/3) \times (1/128) = 1/192$. Det andre aktuelle tilfellet, at Minnie både er homozygot BB og får 7 svarte mus som avkom, har sannsynlighet $(1/3) \times 1 = 1/3$.

Hva blir nå sannsynligheten for at Minnie har genotypen Bb når vi tar hensyn til de 7 svarte avkomsindividerne? Vi regner rett og slett ut hvor stor brøkdel tallet $1/192$ utgjør av summen $(1/192) + (1/3) = 65/192$ av sannsynlighetene for de to tilfellene som er forenelige med svart farge på 7 avkomsindivider. Den endelige sannsynligheten for at Minnie har genotypen Bb , etter at vi har tatt hensyn til observasjonene, blir følgelig $(1/192)/(65/192) = 1/65$. Tilsvarende får vi sannsynligheten $(1/3)/(65/192) = 64/65$ for at Minnie har genotypen BB , gitt opplysningene om de 7 svarte avkomsindividerne.

Ved dette resonnementet er vi kommet frem til nye modifiserte sannsynligheter for de to mulige genotypene BB og Bb for Minnie, ved å bygge på observasjonen av fargen hos avkomsindividerne. Sammenlignet med de opprinnelige sannsynlighetene $1/3$ og $2/3$, har vi rimelig nok endt opp med en mye større sannsynlighet for tilfellet BB , mens sannsynligheten for genotypen Bb er kraftig redusert.

Denne typen problemstillinger er karakteristiske for anvendelser av Bayes' teorem. Vi starter med et sett med *a priori sannsynligheter* (hos oss $1/3$ og $2/3$) for ulike tilstander i naturen, her gitt ved Minnies genotype. Dessuten foreligger det et observasjonsmateriale, slik at vi kan skrive ned sannsynlighetene (her 1 og $1/128$) for de faktiske observasjonsverdiene under forutsetning av at hver av de mulige tilstandene i naturen er korrekte. Ved hjelp av disse sannsynlighetene konverterer vi så de a priori sannsynlighetene til et sett med *a posteriori sannsynligheter* ($64/65$ og $1/65$) for de to mulige tilstandene i naturen.

Ingen vil være uenig i denne typen resonneringer for å finne “inverse” eller a posteriori sannsynligheter. Egentlig dreier det seg bare om en enkel illustrasjon av en vanlig teknikk i sannsynlighetsregning. Uenigheten oppstår dersom det ikke foreligger noen a priori sannsynligheter skrevet opp på grunnlag av en kjent mekanisme som genererer tilfeldige utfall.

Det bayesianske grunnlaget for statistiske slutninger

I eksempelet ovenfor ble alle de aktuelle sannsynlighetene avledet av genetiske modeller. La oss betrakte samme problemstilling med spørsmål om Minnies genotype, men anta at den tilgjengelige forhåndsinformasjonen er litt svakere. Vi går bare ut fra at vi vet at Minnie er svart, ikke at hun var et resultat av en krysning $Bb \times Bb$ mellom to heterozygote individer.

Det er fortsatt klart at Minnie må ha en av genotypene BB og Bb , men hvordan skal vi veie de to mulighetene opp mot hverandre på forhånd? Vi forutsetter naturligvis at det ikke går an å foreta noen nærmere DNA-analyse av Minnies gener! Minnie kan for så vidt stadig være avkom fra en krysning $Bb \times Bb$, men hun kan også være et resultat av krysninger $BB \times BB$, $BB \times Bb$, $BB \times bb$ eller $Bb \times bb$ (hvis vi ikke skiller mellom rekkefølgen av de to angitte foreldregenotypene). Generelt har vi ingen mulighet til å si noe om hvor vanlige disse krysningene skal være i forhold til hverandre. Hvis Minnies foreldre var trukket fra en større populasjon, kunne vi naturligvis gjøre forutsetninger om tilfeldig parring og kanskje bruke regler som Hardy-Weinbergs lov fra populasjonsgenetikken for å beregne sannsynlighetene for ulike foreldregenotyper. Et slikt resonnerement krever imidlertid at vi vet noe om frekvensene til de to allelene B og b i den aktuelle populasjonen, så vi kommer egentlig ikke videre.

En naturlig løsning kan ganske enkelt være å postulere at de to mulige genotypene BB og Bb for Minnie bør ha like stor sannsynlighet, dvs. at begge har verdien $1/2$. I mer uformell språkbruk betyr det at vi i mangel av annen informasjon anser de to tilfellene som “like sannsynlige.” Vi kan nå gjennomføre en regning av samme type som i forrige avsnitt, men med a priori sannsynligheter $1/2$ og $1/2$ for BB og Bb istedenfor $1/3$ og $2/3$. I praksis er dette igjen det samme som å anvende Bayes’ teorem. Etter at vi har tatt hensyn til de 7 observerte svarte avkomsindividene, finner vi a posteriori sannsynligheter for at Minnies genotype skal være henholdsvis BB og Bb som $128/129$ og $1/129$.

Regningen er korrekt ut fra standardreglene for sannsynligheter, men man kan stille seg spørrende til forutsetningene. Det hele avhenger av hva man legger i begrepet “sannsynlighet.” Har vi lov til å regne med a priori verdier $1/2$ og $1/2$ når vi ikke har mer informasjon om Minnies bakgrunn? Det er i slike spørsmål at bayesianernes og frekventistenes veier skiller. En frekventist vil betrakte sannsynligheter generelt som idealiserte verdier tilsvarende relative frekvenser i det lange løp hvis et tilfeldig forsøk gjentas. Når vi ikke vet noe om Minnies opphav, annet enn at hun selv er svart, er det vanskelig å forestille seg hva rimelige gjentakelser skal være i denne situasjonen. Vi kan alltid skaffe til veie mengder av svarte mus ved nye krysninger, men vi vet ikke om de har samme bakgrunn som Minnie. Følgelig må sannsynlighetene for genotypen til Minnie

bare bli stående som ukjente parametere.

En bayesianer vil derimot betrakte sannsynlighetene som uttrykk for den tro man har på de forskjellige mulighetene. I prinsippet bør den som analyserer eksperimentet derfor alltid være i stand til å angi passende a priori sannsynligheter. En zoolog som plutselig stilles overfor en svart mus, har muligens vag forhåndskjennskap til hyppighet av gener som kan påvirke pelsfargen hos gnagere. Er nesten alle mus brune, er det f. eks. lite rimelig å tilordne genotypen BB stor a priori sannsynlighet. På den annen side kan zoologen også anta at Minnie stammer fra et laboratorium ved en annen institusjon, der han mener å huske at nesten alle musene var svarte. Tydeligvis bør a priori sannsynligheten for BB forøyes i dette tilfellet. Er derimot Minnie selv avkom fra en krysning av to heterozygoter Bb , slik vi forutsatte først i regneeksempelet, bør zoologens personlige a priori sannsynligheter falle sammen med de naturlige verdiene gitt ved det opprinnelige resonnementet.

På denne måten kan forskjellige personer som skal analysere samme problemstilling, lett sette opp ulike a priori sannsynligheter. Selv om de studerer samme observasjonsmateriale, vil også konklusjonene gitt ved de a posteriori sannsynlighetene kunne avvike nokså mye fra hverandre. En bayesianer vil understreke at konklusjonene er gyldige for den personen som har formulert de a priori sannsynlighetene, og at det ikke er underlig at mennesker med ulik bakgrunnskunnskap kan ende opp med vidt forskjellige resultater. I motsetning til denne "subjektive bayesianismen" eksisterer det også en retning kalt "objektiv bayesianisme," som prøver å finne a priori sannsynligheter uten noe personlige element som skal influere konklusjonene. Valget av a priori sannsynligheter $1/2$ og $1/2$ for de to mulige genotypene for Minnie kan på denne måten betraktes som et forsøk på å være objektiv.

Hvis man først aksepterer prinsippet med omregning av a priori til a posteriori sannsynligheter i alle situasjoner, har man et kraftig generelt redskap for statistisk analyse. I mange praktiske situasjoner vil de interessante parametrene likevel ha en svært komplisert natur, og utregningene kan bli meget tidkrevende. Ofte vil parameterrommene dekke et kontinuum av aktuelle verdier, og det er nødvendig å foreta integrasjoner i et stort antall dimensjoner for å bestemme a posteriori sannsynlighetsfordelinger. Ser en på slike hindringer som mer tekniske problemer, er det imidlertid mulig å følge samme prinsipielle metode bestandig. Oppskriften med å konvertere a priori sannsynligheter for parametrene til tilsvarende a posteriori verdier ved å inkorporere informasjonen fra dataene kan gjentas flere ganger dersom det blir innsamlet mer data på et senere tidspunkt. I vårt eksempel kan f. eks. Minnie delta i flere krysninger, og vi kan foreta nye omregninger med Bayes' teorem ved å betrakte fargen på de nye avkomsindividene. A posteriori sannsynlighetene som vi endte opp med i vår første regning, kan i så fall benyttes som a priori sannsynligheter i utgangspunktet for de videre utregningene.

I mange problemstillinger vil det inngå flere ukjente parametere enn vi umiddelbart er interessert i å trekke noen konklusjoner om. I så fall må man til slutt summere eller integrere over passende kombinerte a posteriori sannsynligheter

for å nå frem til det endelige resultatet. For en bayesianer er den a posteriori sannsynlighetsfordelingen det fullstendige svaret. For å beskrive vesentlige trekk ved denne fordelingen kan det likevel være naturlig å regne ut f. eks. et intervall som parameteren ligger i med stor sannsynlighet. Hypoteser om parametrene kan veies mot hverandre ved direkte sammenligning av sannsynlighetene for at parametrene skal oppfylle kravene formulert i hypotesene.

Det frekventistiske grunnlaget

Den frekventistiske alternativet til bayesianismen utgjør en mangeartet gruppering. Også frekventister kan ligge i skarp strid seg imellom om grunnprinsippene for statistisk inferens. Betegnelsen frekventister har heller ikke hundre prosent oppslutning. Bayesianeren Lindley (1983) valgte f. eks. å snakke om “Berkeley statistics” som motstykket til bayesiansk statistikk, dels fordi statistikk instituttet ved University of California i Berkeley har vært arnested for mange nyere frekventistiske idéer, og dels fordi filosofen og biskopen George Berkeley i 1736 ble sterkt kritisert av Thomas Bayes for sine angrep på Newton!

Det er et kjernepunkt i de frekventistiske resonnementene at ukjente parametere har faste verdier som observasjonene skal kaste lys over. Man kan gjerne operere med ukjente størrelser som varierer tilfeldig i de aktuelle modellene, men den underliggende sannsynlighetsfordelingen som beskriver denne variasjonen bør være bestemt av faste, muligens ukjente, parametere. Metodene som blir utviklet for statistiske slutninger, vil gi usikre resultater fordi de bygger på observasjoner som følger visse sannsynlighetsfordelinger. Usikkerheten kan naturligvis avhenge av de sanne parameterverdiene, men metodene bør konstrueres slik at de stort sett gir små sannsynligheter for alvorlige feil, uansett hva som måtte være korrekte parameterverdier.

I dette apparatet har sannsynlighetsbegrepet bare en plass som abstraksjon av begrepet relative frekvenser, som kan observeres i naturen ved et stort antall gjentagelser av et tilfeldig forsøk. I modellene vil ukjente sannsynligheter ofte utgjøre viktige parametere. Disse sannsynlighetene gir ikke primært uttrykk for personlig tro på at bestemte begivenheter skal finne sted, men gjenspeiler snarere en korrekt beskrivelse av naturen som vi bare oppnår et ufullstendig bilde av. Vårt enkle eksempel er litt annerledes siden Minnies genotype blir betraktet som aktuell parameter, enda denne størrelsen bare kan ha to mulige tilstander BB og Bb .

I praksis har utviklingen av frekventistiske metoder for statistisk inferens gått mange veier. En retning fremhever betydningen av å sammenligne sannsynlighetene for det observasjonsresultatet som vi faktisk har fått, under forskjellige aktuelle verdier for de ukjente parametrene (Edwards, 1972). I denne sammenhengen blir hver sannsynlighet ansett som en “likelihood” (på norsk kanskje “rimelighetsverdi”). I vårt eksempel vil det være snakk om verdiene 1 hvis genotypen i virkeligheten er BB og $(1/2)^7 = 1/128$ hvis genotypen er Bb . En 128 ganger så stor “likelihood” gir tydeligvis sterke holdepunkter for at genotypen BB er den korrekte, selv om konklusjonen kan være feil. I andre, mer kompliserte situasjoner er det ikke alltid like lett å følge samme prinsipp for

statistiske slutninger.

En annen retning konsentrerer seg om å finne metoder som i alminnelighet gir små sannsynligheter for feilaktige konklusjoner, enten det dreier seg om testing av hypoteser eller utregning av anslag av ukjente størrelser. Dette prinsippet ligger bak velkjente statistiske metoder som signifikanstesting eller intervallestimering. Den generelle tankegangen går tilbake til mye av det grunnleggende arbeidet som ble utført av R. A. Fisher i 1920- og 1930-årene med utvikling av statistiske metoder, og til de noe avvikende standpunktene representert av arbeidene til Jerzy Neyman og Egon Pearson bare litt etter Fishers pionerinnsats.

Problemet med Minnies genotype er nesten for enkelt til å bli tilpasset denne tankegangen. Dersom det dukker opp et eller flere brune avkomsindivider i kryssningen mellom Minnie og en brun hannmus, er det straks avslørt at den sanne genotypen til Minnie er Bb og ikke BB . Vi kan derfor lage oss en enkel statistisk test som for det første bygger på denne betraktningen, og som ellers konkluderer med at Minnies genotype er BB dersom alle 7 avkomsindividene er svarte. Hvis den korrekte genotypen er Bb , vil denne prosedyren gi feil konklusjon med sannsynligheten $(1/2)^7 = 1/128$ og riktig konklusjon med sannsynlighet $127/128$. Hvis den korrekte genotypen er BB , er sannsynligheten for riktig konklusjon lik 1. Ut fra sedvanlige kriterier for feilsannsynligheter har denne testmetoden meget bra egenskaper, selv om en observasjon tilsvarende 7 svarte mus naturligvis kan føre til gal konklusjon her også.

Argumenter for bayesiansk tankegang

Den bayesianske tankegangen gjør hele apparatet for statistisk inferens prinsipielt mye enklere enn frekventistisk tankegang. All usikkerhet om forhold i naturen blir uttrykt på samme måte ved hjelp av sannsynligheter. Det kan riktignok være nødvendig for en person som vil bruke de bayesianske metodene å øve seg opp i fastsettelse av fornuftige personlige sannsynligheter. Den påfølgende matematiske behandlingen vil likevel alltid følge de samme prinsippene. På denne måten kan man opprettholde en enhetlig behandling av problemstillinger der usikkerhetsmomentet i utgangspunktet er av svært ulik karakter. Sammenfatning av usikkerhet i forbindelse med langsiktig global klimautvikling, med mer eller mindre relevante data fra forskjellige kilder i tillegg til personlige standpunkter uttrykt av angivelige autoriteter, skal i teorien behandles etter de samme prinsippene som f. eks. analyse av et sett nøyaktige målinger av gravitasjonskonstanten. En bayesianer vil naturligvis innrømme at det er mye vanskeligere å stille opp de riktige sannsynlighetsfordelingene i den første tilfellet, men det rokker ikke ved de generelle retningslinjene.

En annen fordel ved den bayesianske tankegangen er at man lett kan kombinere informasjon av helt ulik natur for å belyse *samme* problem. Så fremt informasjonen enten kan uttrykkes i en subjektiv a priori sannsynlighetsfordeling, eller det er mulig å skrive opp sannsynligheten for det man har fått kjennskap til, gitt de mulige sanne tilstandene i naturen, vil informasjonen kunne bygges inn i den a posteriori sannsynlighetsfordelingen. Hvis det er aktuelt, er det mulig å oppdatere denne fordelingen trinnvis, dersom informasjonen blir tilgjengelig

i ulike stadier etter hverandre. Frekventistiske metoder må i mange tilfeller gjennomgås fra bunnen dersom det skulle dukke opp nye observasjoner.

Bayesianerne understreker gjerne at det enhetlige preget ved deres egen tankegang avviker sterkt fra “ad hoc-metodene” som ofte blir utviklet i frekventistisk statistikk. Mange av metodene der blir kritisert fordi de bygger på forhold også ved andre mulige sett med observasjonsverdier enn akkurat det som virkelig er registrert. Dette skjer fordi frekventistisk statistikk ofte legger større vekt på egenskapene ved selve *metoden* enn at man skal kunne måle hvor sterk konklusjonen blir med det gitte datasettet. Bayesianske metoder krever derimot bare at man kan oppgi sannsynlighetene for akkurat de observerte verdiene under ulike muligheter for parametrene. Denne egenskapen deler de for så vidt også med den rendyrkede “likelihood”-metoden i frekventistisk tankegang.

I visse statistiske problemstillinger fastlegger man ikke på forhånd hvor mange observasjoner som skal tas, men lar det bli bestemt av resultatene som kommer underveis. I slike situasjoner vil mange frekventistiske metoder for analyse av observasjonsrekken avhenge vesentlig av regelen for når man skal stoppe. Det har vært ansett som et argument for bayesianske metoder og “likelihood”-metoder at konklusjonen her bare avhenger av det bestemte settet med observasjoner som man har fått, uansett hvilken stopperegulering som er anvendt.

Begrep som P -verdi ved en signifikanstest og konfidenskoeffisient for et konfidensintervall feiltolkes ofte av ukyndige brukere av faget. En vanlig misforståelse er at P -verdien for et observasjonsresultat er sannsynligheten for at nullhypotesen er riktig. Tilsvarende blir konfidenskoeffisienten fortolket som sannsynligheten for at den virkelige parameterverdien skal ligge i det utregnede intervallet. I realiteten er slike frekventistiske begrep knyttet til *metodene* som er anvendt, og sannsynlighetene refererer seg til andre mulige datasett som kunne ha vært observert. De direkte, litt naive forestillingene hos brukerne stemmer imidlertid godt overens med bayesianske idéer. Av og til kan det virke som om bayesianerne utnytter dette poenget som det siste avgjørende moment til sin egen fordel!

Kritikk av bayesianske idéer

Den viktigste motforestillingen mange statistikere har mot generell bruk av bayesianske metoder henger sammen med valget av a priori sannsynligheter. For det første er det vanskelig i mange situasjoner å angi sannsynligheter som skal gjenspeile ens forhåndsinnstilling til ukjente størrelser. Bayesianerne har nedlagt mye arbeid i å vise teoretisk hvordan sannsynligheter bør velges hvis ens tiltro til de forskjellige muligheter oppfyller visse logiske krav. Man støtter seg også til hjelpeprosedyrer slik at de aktuelle mulighetene kan sammenlignes med andre begivenheter som det kan være lettere å tallfeste tiltroen til. Likevel vil problemet med valg av passende a priori sannsynligheter lett virke som en hemsko ved bruk av bayesianske metoder.

Selv om man klarer å stille opp et sett med a priori sannsynligheter som er tilfredsstillende for en selv, vil man ofte ha store problemer med å få andre til å akseptere ens valg. A priori sannsynlighetenes subjektive natur gjør det vanskelig å argumentere senere overfor andre når det gjelder konklusjonene som

den statistiske metoden leder til. I noen situasjoner kan man prøve med a priori sannsynlighetsfordelinger som er tilnærmet “ikke-informative,” i analogi med sannsynlighetene $1/2$ i vårt tidligere genetiske eksempel. Det er imidlertid problematisk å definere slike fordelinger som tilsvarer fullstendig uvisshet i situasjoner der parameterrommet har en bestemt struktur, som f. eks. med reelle tall. Foretar man en enkel transformasjon av parametrene, vil den tilsynelatende uvissheten vanligvis gå tapt.

Som svar på denne kritikken vil bayesianerne ofte påpeke at valget av a priori sannsynligheter har liten innflytelse på de endelige a posteriori sannsynlighetene når observasjonene inneholder svært mye informasjon. I mange standardmodeller kan det faktisk være likegyldig om man bruker bayesianske eller frekventistiske metoder så lenge datamaterialet er svært stort. Men ofte er det nettopp tilfellene med begrenset informasjon i de tilgjengelige datamengdene som er interessante fra en anvendt synsvinkel. Det kan i mange tilfeller være vanskelig å komme utenom kritikken her, selv om bayesianerne har rett når de understreker at også frekventistene bygger inn forhåndskunnskap i en annen forstand ved selve modelleringen.

Mer generelt får en lett følelsen av at bayesianske metoder blander sammen bevisgrunnlag av forskjellig natur for bestemte hypoteser. I situasjoner der en prøver å analysere et datasett etter en prosedyre som er tilnærmet objektiv, virker det underlig å trekke inn løs forhåndsinformasjon fra helt andre kilder. Det kan i og for seg være naturlig å tenke på slike opplysninger når hver enkelt skal gjøre seg opp sin personlige oppfatning senere. Men det er ikke dermed gitt at disse opplysningene bør betraktes i en nøktern analyse av et nytt datasett, spesielt når resultatene skal kunne aksepteres av folk med svært ulik forhåndsinnstilling.

Det har hittil vært en vesentlig ulempe ved de bayesianske metodene at de a posteriori fordelingene mange ganger er så vanskelige å beregne numerisk. Hvis den a priori fordelingen, skrevet opp som matematisk uttrykk, tilhører bestemte klasser, finnes det i visse statistiske standardsituasjoner teoretiske resultater som viser at den a posteriori fordelingen også må tilhøre samme klasse. Men utenfor disse situasjonene har det i mange praktiske tilfeller vært nærmest umulig å regne ut a posteriori fordelingene. Ofte har det vært tale om numerisk integrasjon med så mange variable at det hele ble ugjennomførbart. Med den senere tids utvikling av regnekraft er imidlertid bildet blitt noe annerledes.

Utviklingen de siste årene

Tradisjonelt stoppet debatten mellom bayesianere og frekventister på dette stadiet. Det virket i mange år som om det var umulig å komme videre og skape noen bevegelse i frontene. Bayesianerne gjentok gjerne de samme fundamentale idéene (Lindley, 1975, 1978, 1983), og frekventistene beskrev praktiske problemer som vanskelig lot seg løse med de tilgjengelige bayesianske angrepsmåtene (Efron, 1986). I de siste årene har man likevel kunne iaktta en oppmykning av konflikten. Det skyldes dels at mer generelle problemstillinger er blitt trukket inn i statistikkfagets domene, noe som igjen har hatt til følge at andre grunnleggende

betraktningmåter også har vunnet innpass. Dessuten er det blitt mulig å anvende dataintensive metoder i en helt annen utstrekning enn før, og denne utviklingen har også åpnet for mer prinsipielle forandringer i statistikernes innstilling til ulike angrepsmåter.

Tross grunnleggende problemer er det tydelig at bayesianske metoder har vunnet større innpass på forskjellige anvendelsesområder. Breslow (1990) gav en oversikt over bruk av bayesianske metoder i biostatistikk, og han omtaler en rekke anvendelser der bayesiansk modellering ser ut til å føre lenger enn tradisjonell frekventistisk tankegang. Det gjelder spesielt situasjoner der statistisk inferens kan gjennomføres på flere nivåer, f.eks. ved analyse av insidensrater for kreft i geografiske områder som støter opp mot hverandre. Her kan det både foretas en lokal analyse, og dessuten en analyse av alle områdene under ett, slik at det også tas hensyn til variasjonen mellom områder. Mange av disse modellene er imidlertid ikke bayesianske i streng forstand. I dette eksempelet kan man f.eks. innføre en sannsynlighetsfordeling for de underliggende parametrene i et vilkårlig geografisk område. Ofte er denne fordelingen uttrykt ved nye “overordnede parametre” som gjelder hele studien, og disse parametrene blir så anslått ut fra dataene. Dette er snarere metoder som faller inn under kategorien “empirisk Bayes-statistikk,” der man ikke opererer med ekte a priori sannsynligheter som uttrykker forhåndskjennskap. Randomiserte kliniske forsøk er et annet viktig område i medisinsk statistikk der bayesianske metoder kan vise seg å bli verdifulle (Spiegelhalter *et al.*, 1994).

Mange av problemstillingene som ser ut til å egne seg for bayesianske metoder, har mer karakter av beslutningsproblemer enn rene vitenskapelige undersøkelser. Overgangen kan likevel være glidende. Det kan spesielt gjelde problemer av økonomisk natur, eller situasjoner der det er mulig å kvantifisere konsekvensene av ulike handlinger. Hvis et organ blir presset til å ta beslutninger med alvorlige potensielle konsekvenser, og samtidig skal ta hensyn til observasjonsmaterialer med stor usikkerhet, er det naturlig at konklusjonene formes på en annen måte enn i grunnlagsforskning, der resultatene bør ha stor allmenngyldighet. Ønsket om å bygge på absolutt all tilgjengelig informasjon blir betraktelig større i en beslutningssituasjon. Samtidig behøver en ikke i samme grad bekymre seg fordi ens egen a priori sannsynlighetsfordeling kanskje ikke stemmer med sannsynligheter som andre ville formulere.

Det har lenge vært akseptert at bayesianske metoder kan være berettiget ved beslutningsproblemer av denne typen. Det kan f.eks. gjelde risikovurderinger ved analyse av kompliserte systemer, der kjennskapet til de ulike komponentene er av forskjellige slag. Mange frekventister har også sett behovet for bayesianske metoder som et supplement i mer vitenskapelige studier med hypoteseprøving, men de numeriske vanskelighetene har ofte utgjort et uoverstigelig hinder for innføring av realistiske modeller. Spesielle teknikker har vært utviklet for numerisk beregning av a posteriori sannsynlighetsfordelinger. De siste årene har sett store forandringer på dette området med innføring av såkalte “Markovkjede-Monte Carlo-metoder” for simulering av aktuelle sannsynligheter (Besag *et al.*, 1995). Teknikkene utnytter betingede sannsynlighetsfordelinger

mellom noen av de relevante variablene til simulering av mer kompliserte simultane sannsynlighetsfordelinger der det inngår mange variable. Resultatet er en effektiv beregning av de a posteriori sannsynlighetsfordelingene med stor grad av nøyaktighet, også i uoversiktlige situasjoner som ellers ville vært umulige å behandle.

Disse metodene har bl. a. fått betydning i sammenheng med en fornyet interesse for hypoteseprøving i bayesiansk forstand (Kass og Raftery, 1995). Prinsipielt sammenlignes tiltroen til ulike hypoteser ved å se på forholdet mellom de a posteriori sannsynlighetene for hypotesene. Det blir understreket at denne fremgangsmåten egner seg til å samle empirisk bevismateriale *til fordel for* bestemte hypoteser, i motsetning til frekventistisk hypoteseprøving, som vesentlig tar sikte på å *forkaste* eller eventuelt ikke forkaste bestemte hypoteser. Utviklingen på dette området tar tydeligvis lite hensyn til generelle idéer fra vitenskapsteorien. Øyensynlig har dette vært tilfelle i lengre tid for folk som har syslet med grunnlagsproblemer for statistisk inferens (Good, 1988).

De bayesianske idéene har også brakt noe nytt inn i forhold til tradisjonelle statistiske problemstillinger ved å gjøre det mulig å ta hensyn til usikkerheten som ligger i selve modellvalget (Kass og Raftery, 1995). Hvis det kan stilles opp a priori sannsynligheter for ulike modeller for databeskrivelsen, kan disse verdiene på vanlig måte konverteres til a posteriori sannsynligheter når dataene er kjent. Som en anvendelse litt på siden av de statistiske inferensproblemene står bayesianske beregningsmåter benyttet ved modellering av usikkerhet i ekspert-systemer (Spiegelhalter *et al.*, 1993). Slike systemer skal f. eks. lede til diagnose av bestemte sykdommer på grunnlag av mange usikre observasjoner på en pasient.

Utviklingen innenfor bayesiansk statistikk de siste årene har utvilsomt skapt større generell entusiasme for grunnleggende bayesianske idéer. Samtidig kan en registrere en øket interesse iallfall for utprøving av bayesianske metoder på stadig nye anvendelsesområder. Det er usikkert om tendensen vil holde seg, og det er vanskelig å angi noen personlig sannsynlighet for at den bayesianske retningen skal bli dominerende. Lindley (1977) spådde at vi kommer til å gå inn i et bayesiansk århundre, og i denne forbindelsen uttalte frekventisten Efron (1978) at hans personlige sannsynlighet for denne begivenheten var 0.15! Det er mulig at sannsynligheten har øket i mellomtiden. Det virker likevel mer rimelig å tro at det i fremtiden vil etableres en bedre balanse mellom bayesianske og frekventistiske angrepsmåter, og at begge retningene vil finne sin plass i idégrunlaget for statistikkfaget.

Referanser

- Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc. London* **53**, 370–418. Gjengitt i G. A. Barnard: Studies in the history of probability and statistics. IX. *Biometrika* **45** (1958), 293–315.

- Besag, J., Green, P., Higdon, D., Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10**, 3–66.
- Breslow, N. (1990). Biostatistics and Bayes. *Statist. Sci.* **5**, 269–298.
- Edwards, A. W. F. (1972). *Likelihood*. Cambridge: Cambridge University Press.
- Efron, B. (1978). Controversies in the foundations of statistics. *Amer. Math. Month.* **85**, 231–246.
- Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.* **40**, 1–5.
- Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh: Oliver and Boyd.
- Good, I. J. (1988). The interface between statistics and philosophy of science. *Statist. Sci.* **3**, 386–412.
- Kass, R. E., Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773–795.
- Lindley, D. V. (1975). The future of statistics – a Bayesian 21st century. *Supp. Adv. Appl. Prob.* **7**, 106–115.
- Lindley, D. V. (1978). The Bayesian approach. *Scand. J. Statist.* **5**, 1–26.
- Lindley, D. V. (1983). Theory and practice of Bayesian statistics. *The Statistician* **32**, 1–11.
- Natvig, B. (1989). På tide med Bayes-analyse i Norge nå? *Tilfeldig Gang* **6** (3), 7–13.
- Norberg, R. (1991). Bayes-metodikkens plass i statistiske beslutningsproblemer. Hva er subjektivt, og hvem er subjektivister? *Tilfeldig Gang* **8** (1), 9–16.
- Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statist. Sci.* **8**, 219–283.
- Spiegelhalter, D. J., Freedman, L. S., Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *J. Roy. Statist. Soc. A* **157**, 357–416.
- Sverdrup, E. (1990). Hva er Bayes prinsipp? Hvem er Bayesianere? *Tilfeldig Gang* **7** (1), 5–13.