

I PRAKSIS

FREKVENTISME, BAYES OG LIKELIHOODPRINSIPPET

Av Emil Aas Stoltenberg

1

Jeg ser at du smiler og jeg tenker at du er glad. På bakgrunn av en observasjon (et smil) gjør jeg en slutning om hvordan jeg tror verden er (du er glad). Få ting er like grunnleggende for oss mennesker som denne typen induktive slutninger. Vi gjør slike slutninger nesten hele tiden hver dag, og all empirisk vitenskap innebærer å slutte fra observasjoner til hvordan man tror verden er. Faget statistikk består av matematiske formaliseringer av slike slutningsprosesser. Faget er normativt i den forstand at disse formaliseringene ikke forsøker å beskrive hvordan vi mennesker faktisk foretar slike slutninger, men hvordan vi rasjonelt *bør* la data (altså observasjoner) forme vår oppfatning om verden, og det er her, ved dette 'bør', at uenigheten jeg vil ta for meg i denne teksten oppstår. I følge Hacking (1965) er statistikkens grunnlagsproblem 'å formulere et sett prinsipper som medfører gyldigheten til alle korrekte statistiske slutninger.'¹ Prinsippene bør fremstå som åpenbare og de statistiske metodene skal ikke kunne lede til motsigelser. I denne teksten vil jeg presentere et slikt prinsipp og to forskjellige tilnærminger til statistikk.

I tekstens første del vil jeg presentere den vanligste måten å tilnærme seg statistikk på, nemlig den frekventistiske. At jeg starter med denne er ikke fordi den er lettest å forstå, det er den ikke, men fordi vi senere skal se hvordan den bryter med noen foreslåtte prinsipper for statistiske slutninger. I tekstens andre del presenterer jeg et slikt prinsipp, kjent som likelihoodprinsippet. Det faktum at den frekventistiske tilnærmingen til statistikk bryter med likelihoodprinsippet har ledet mange statistikere til det som kalles bayesiansk statistikk. I tekstens tredje del presenterer jeg den bayesianske tilnærmingen til statistikk og diskuterer noen av dens viktigste utfordringer. Jeg holder matematikken til et minimum. Derfor trenger jeg et eksempel

for å illustrere idéene og argumentene jeg presenterer, og siden mange av disse argumentene er ganske vanskelige bør dette eksempelet være enkelt. I likhet med utallige andre tekster om statistikk og sannsynlighetslære starter jeg derfor med myntkastet.

2

Vi er alle enige om at i det jeg legger et kronestykke på tommel og pekefinger og gjør meg klar til å flippe kronestykket opp i luften kan jeg ikke være sikker på om kronestykket litt senere vil vise kron eller mynt. Vi oppfatter utfallet som tilfeldig. For Laplaces demon, som kjenner hvert av universets grunnleggende bestanddeler posisjon og hastighet vil det kanskje være mulig å forutsi utfallet, men for oss mennesker er det ikke det. Om det vi kaller tilfeldighet er en iboende egenskap ved naturen, eller om det er vi mennesker som oppfatter ting som tilfeldig fordi det er grenser for hva vi kan vite noe om, er et spørsmål om hva sannsynlighet *er*. Er sannsynlighet en egenskap knyttet til mynten og myntkastet, eller er sannsynlighet bare et tall vi bruker for å uttrykke vår usikkerhet forbundet med et fremtidig utfall eller en ukjent størrelse? Heldigvis er det slik at for den delen av matematikken som kalles sannsynlighetsregning har disse spørsmålene ingenting² å si. Når det gjelder sannsynlighetsregning som en formell matematisk disiplin er det altså ingen uenighet. Det er først når vi ønsker å bruke sannsynlighetsregningen til å lære noe om verden at uenighetene oppstår, og på denne måten er spørsmålet om hva sannsynlighet er tett knyttet til hvordan man mener at statistikk bør bedrives. Vi skal komme tilbake til dette.

Inntil videre, la oss flippe det samme kronestykket tolv ganger og notere utfallene

M M M M M K M K M M M K

Uansett hva man måtte mene at sannsynlighet uttryk-ker er det åpenbart at disse tolv myntkastene lærer oss noe. Med de tolv utfallene bør vi være i bedre stand til å gjette på utfallet av det trettende kastet eller vurdere om vi har å gjøre med et rettferdig kronestykke, enn uten disse tolv myntkastene. I den neste delen av teksten skal jeg beskrive hvordan den frekventistiske statistikken tar oss fra disse tolv datapunktene til statistiske slutninger om verden.

3

Frekventistisk statistikk er, som navnet tilsier, tett knyt-tet til det syn at sannsynlighet uttrykker hypotetiske fre-kvenser i det lange løp. Det betyr at dersom vi flipper kronestykket vårt uendelig mange ganger vil andelen mynt konvergere mot et tall, og dette tallet er det vi kal-ler sannsynligheten for mynt (og 1 minus dette tallet er sannsynligheten for kron). I følge dette synet er denne sannsynligheten en egenskap ved mynten på samme måte som vekt og omkrets er egenskaper ved denne mynten, lengde er en egenskap ved en planke og så videre. Og på samme måte som vi bruker en tommestokk for å måle lengden til en planke, bruker vi et endelig antall myntkast for å måle egenskapen 'sannsynlighet' til et kronestykke.

Den frekventistiske (eller objektive) måten å betrakte sannsynlighet på har implikasjoner for hva slags sannsyn-lighetsytringer som er meningsfulle. Siden kronestykkets tilbøyelighet til å vise mynt er en konstant størrelse, er det med den frekventistiske tolkingen av sannsynlighet meningsløst å hevde at denne størrelsen ligger i intervallet $1/3$ til $2/3$ med en viss sannsynlighet. Enten ligger denne størrelsen i dette intervallet, eller så gjør den det ikke. Ytringen er altså enten sann eller usann, har sannsynlighet 1 eller 0 og ikke noe i mellom. Dette medfører at vi, i prinsippet, ikke kan si noe med en viss usikkerhet om kronestykkets tilbøyelighet til å vise mynt. Det vi kan si noe om er sannsynligheten for at et kronestykke viser mynt i tolv av tolv kast dersom vi, for eksempel, antar at kronestykket er rettferdig. Denne sannsynligheten er $1/4096$, altså ganske lav, så om vi observerer tolv mynt i tolv kast har vi liten grunn til å tro at mynten er rettferdig. Det viktige å merke seg her er at sannsynlighetsytringen knytter seg til utfallene av myntkast under en gitt antagelse om kronestykkets beskaffenhet, og ikke til kronestykkets beskaffenhet som sådan.

3.1

La oss si at du vil finne ut om den mynten vi har kastet tolv ganger er rettferdig. Med 'rettferdig' mener jeg at sannsynligheten for kron er den samme som sannsynligheten for mynt, altså en halv. For å finne ut av om kronestyk-ket er rettferdig starter den frekventistiske statistikerer med å formulere en såkalt nullhypotese. Den naturlige nullhypotesen i dette tilfellet er at det er like sannsynlig å kaste kron som mynt.

Med denne hypotesen ferdig formulert må statisti-keren stille seg følgende spørsmål: Hvilke observasjoner vil få meg til å slutte å tro på nullhypotesen? Svaret på dette spørsmålet kalles en forkastningsregel. Det er en re-gel som forteller statistikerer om hun skal fortsette å tro på nullhypotesen, eller forkaste nullhypotesen til fordel for en alternativ hypotese. Den alternative hypotesen er gjerne nektelsen av nullhypotesen, 'kronestykket er ikke rettferdig'. Et eksempel på en forkastningsregel er 'jeg vil ikke lenger tro på nullhypotesen hvis det er tre eller færre kron eller mynt blant tolv myntkast'. Med denne forkastningsregelen ville de tolv myntkastene ovenfor, som jo ga 9 mynt og 3 kron, lede oss til å ikke lenger tro at vi har å gjøre med et rettferdig kronestykke, med andre ord så forkaster vi nullhypotesen.

Legg merke til at selv om sannsynligheten for tre el-ler færre kron eller mynt i tolv kast med et rettferdig kro-nestykke ikke er veldig stor, så vil det skje ganske ofte. Sannsynligheten for at det skjer er 14.6 prosent. Det betyr at med vår forkastningsregel er sannsynligheten for at vi feilaktig forkaster nullhypotesen 14.6 prosent. I stedet for å velge seg noe som intuitivt fremstår som en fornuftig forkastningsregel, pleier statistikerer å velge seg en forkast ningsregel som er slik at sannsynligheten for å forkaste en sann nullhypotese er mindre enn eller lik en gitt terskel-verdi, ofte 5 prosent. Man regner altså ut på forhånd hvil-ken x som er slik at sannsynligheten for 'x eller færre kron eller mynt blant tolv mynkast' er mindre enn eller lik den valgte terskelverdien når kronestykket er rettferdig. Valget av terskelverdi avhenger gjerne av hvor ille det er å feilaktig slutte å tro på nullhypotesen.

Men det er en annen måte vi kan ta feil på også. Vi kan fortsette tro at kronestykket er rettferdig selv om det i virkeligheten ikke er det. Sannsynlighetene (i flertall!) for dette er ikke like enkle å regne ut fordi det er veldig mange måter (ikke-tellbart³ mange måter) kronestykket kan være urettferdig på. Av denne grunn er det vanlig å rigge hypo-tesene slik at det er verre å begå den første feilen (forkaste en sann nullhypotese) enn den andre (beholde en usann nullhypotese), nettopp fordi vi kan kontrollere sannsyn-

ligheten for den første feilen ved valg av forkastningsregel. Samtidig som vi vil vokte oss for å forkaste en sann nullhypotese, er det klart at dersom nullhypotesen er usann, håper vi at vi forkaster den. Sagt på en annen måte så ønsker vi å maksimere sannsynligheten for å forkaste en usann nullhypotese under den betingelse at sannsynligheten for å forkaste en sann nullhypotese ikke overstiger vår valgte terskelverdi. En forkastningsregels evne til å oppdage at nullhypotesen er usann kalles forkastningsregelens styrke, og i statistikken finnes det teoremer som forteller en hvordan man skal konstruere forkastningsregler som er så sterke som mulig samtidig som de ikke overstiger terskelverdien dersom nullhypotesen er sann.

Dette lynkurset i den frekventistiske tilnærmingen til hypotesetesting kan oppsummeres i følgende punkter.

- (1) Formuler en nullhypotese.
- (2) Velg en terskelverdi.
- (3) Velg en forkastningsregel som maksimerer styrken.
- (4) Sjekk hva data forteller deg.

4

En av de viktigste innvendingene mot den frekventistiske tilnærmingen til statistikk har sitt utspring i det som kalles likelihoodprinsippet. Ifølge dette prinsippet skal støtten de observerte data gir til konkurrerende hypoteser kun avhenge av hvor sannsynlig de observerte data er under hver av de konkurrerende hypotesene. Det var en litt tung setning, men dersom du leser den en gang til og forstår den vil du kanskje tenke, 'ja, hva ellers?'. Nettopp fordi dette er lett å akseptere holdes det av mange opp som et prinsipp.

La oss se hva dette betyr for de tolv myntkastene våre og den frekventistiske fremgangsmåten vi ovenfor brukte for å teste om kronestykket var rettfærdig. De to konkurrerende hypotesene er 'sannsynligheten for kron er en halv' (nullhypotesen) og 'sannsynligheten for kron er ikke en halv'. Anta at vi krever at sannsynligheten for å feilaktig forkaste nullhypotesen skal være mindre enn 5 prosent. Forkastningsregelen, der vi forkaster nullhypotesen dersom vi observerer 2 eller færre kron eller mynt, tilfredsstiller denne terskelverdien.

I eksempelet vårt observerte vi tre kron og ni mynt, så våre data leder til at vi beholder nullhypotesen. Men (dette er det viktige), merk at sannsynligheten som i vårt tilfelle gir støtte til nullhypotesen om at kronestykket er rettfærdig er sannsynligheten for tre, fire, fem, seks, syv, åtte, ni eller ti kron eller mynt. Men vi observerte tre kron, ikke flere! Dette betyr at vårt valg om å beholde nullhypotesen, eller med andre ord, støtten data ga til nullhypotesen, er avhen-

gig av data som kunne vært observert, men ikke var det.

For å understreke det kontraintuitive med denne praksisen er det fristende å parafrasere Jeffreys⁴: Vi har beholdt en nullhypotese som kanskje er usann fordi den har forutsett mulige utfall som ikke ble observert. På denne måten bryter den frekventistiske metoden vi brukte ovenfor med likelihood-prinsippet. Vi skal nå se på et klassisk eksempel som kanskje enda tydeligere får frem det kontraintuitive som brudd på likelihoodprinsippet kan lede til.

4.1

Ovenfor bestemte vi oss for å flippe det samme kronestykket tolv ganger. Vi noterte utfallene og fikk tre kron og ni mynt. Anta at vi i stedet for å på forhånd bestemme oss for å flippe kronestykket tolv ganger, så ble vi enige om å flippe kronestykket helt til vi observerte vår tredje kron. Titt igjen på de tolv myntkastene vi har notert ned utfallene av ovenfor og legg merke til at disse like gjerne kunne kommet i stand på denne andre måten. Det tolvte kastet landet jo kron. Vi ønsker igjen å teste om kronestykket er rettfærdig og har igjen valgt forkastningsregelen slik at vi kun forkaster en sann nullhypotese med mindre enn 5 prosent sannsynlighet.

Nå kommer det underlige. Dersom vi på forhånd har bestemt oss for å flippe kronestykket tolv ganger beholder vi nullhypotesen. Men, hvis vi derimot har bestemt oss for å kaste inntil vi har observert tre kron, da forkaster vi nullhypotesen. Her leder altså de samme observasjonene til forskjellige konklusjoner avhengig av hvordan vi har bestemt oss for å stoppe flippingen av kronestykket (ved det tolvte kastet eller ved tredje kron). Intuitivt virker det urimelig fordi vi tross alt har å gjøre med det samme kronestykket, den samme ukjente størrelsen vi ønsker å si noe om, og tre av tolv er tre av tolv uansett hvordan disse tre av tolv er kommet i stand.

Matematisk er det derimot ikke vanskelig å forstå hvorfor vi ledes til motstridende konklusjoner. Grunnen til at de samme observasjonene gir forskjellige konklusjoner er at hvilke data vi kunne observert er forskjellige avhengig av om vi bestemte oss for å flippe kronestykket tolv ganger eller stoppe ved tredje kron. I det første tilfellet kunne antall kron vært hvilket som helst tall mellom null og tolv. I det andre tilfellet vet vi at det siste myntkastet må være kron, og at det blant de resterende vil befinne seg to andre kron. Det er denne forskjellen i mulige utfall som leder til forskjellige konklusjoner. Men at noe er mulig rokker ikke ved det faktum at kronestykket vårt viste kron tre ganger i tolv kast. At noe som er mulig, men ikke har skjedd, skal ha noe å si virker rart.

Likelihoodprinsippet formaliserer denne intuisjonen og gjør til prinsipp at data vi hypotetisk sett kunne ha observert, men som vi ikke har observert, skal være irrelevante for konklusjonene vi trekker fra data. Tar man likelihoodprinsippet alvorlig er det vanskelig å være frekventist, fordi det er ubehagelig å drive med slutningsprosedyrer som bryter med dette prinsippet. Når det er sagt, er det ikke alle statistikere som mener likelihoodprinsippet er så intuitivt som dets tilhengere hevder. Derfor var det banebrytende da Birnbaum (1962) viste at likelihoodprinsippet er ekvivalent med sammenstillingen av to andre prinsipper som for mange er lettere å akseptere og, ikke minst, er grunnleggende for mye frekventistisk statistikk. Birnbaum viste at dersom man aksepterer det som kalles suffisiensprinsippet og prinsippet om betingethet, må man med nødvendighet akseptere likelihoodprinsippet. I stedet for de presise definisjonene vil jeg nå gi et par eksempler på hva disse to prinsippene sier.

Suffisiensprinsippet først. Se for deg at du og en venn av deg er interessert i gjennomsnittshøyden på basketballspillere i Norge. Du trekker et tilfeldig utvalg basketballspillere, vennen din trekker et annet tilfeldig utvalg basketballspillere. Basketballspillerne i ditt utvalg er ikke de samme som i din venn sitt utvalg, deres datapunkter er altså ikke de samme. Men, anta at gjennomsnittshøyden av ditt utvalg er det samme som gjennomsnittshøyden i din venn sitt utvalg. Suffisiensprinsippet sier da at dere skal komme til den samme konklusjonen om den underliggende sanne gjennomsnittshøyden i populasjonen av basketballspillere i Norge.

Et eksempel til: Du og din venn bestemmer dere for å kaste den samme mynten tolv ganger hver. Du får rekken av myntkast over. Din venn kaster mynt, kron, kron, kron, og deretter åtte mynt. Dere har ikke de samme observasjonene, deres rekker av kast er forskjellige, men begge har observert tre av tolv kron. Da sier suffisiensprinsippet at dere skal mene det samme om kronestykkets tilbøyelighet til å lande med kron opp. Nå, prinsippet om betingethet. Se for deg at Kunnskapsdepartementet vil teste ut bruk av noe nytt undervisningsmaterieell i skolen, iPad til alle elever for eksempel. Både Oslo og Bergen har lyst til å stille med forsøksskoler. Kunnskapsminister Torbjørn Røe Isaksen vil ikke bli beskyldt for å favorisere den ene byen og bestemmer seg derfor for å avgjøre hvem som blir tildelt eksperimentet ved et myntkast. Bergen vinner, mottar iPadene, gjennomfører eksperimentet og sender sine data tilbake til Kunnskapsdepartementet. Prinsippet om betingethet forteller Røe Isaksen at han ikke skal instruere analyseavdelingen i Kunnskapsdepartementet om å

ta med i sine beregninger at iPad-eksperimentet, som ble gjennomført i Bergen, kunne ha vært gjennomført i Oslo. Med andre ord skal myntkastet utført på kunnskapsministerens kontor være irrelevant for de statistiske slutningene. Det motsatte virker urimelig.

Et siste eksempel: Du føler deg litt dårlig, er redd du har feber og vil ta temperaturen. Siden det er kjipt å gå glipp av exphil-forelesningen om Peter Singer er din nullhypotese at du er frisk. Du vil kontrollere sannsynligheten for å feilaktig bli i senga når du egentlig er frisk. Dessverre finner du ikke ditt splitter nye termometer, og må ta til takke med det gamle. Prinsippet om betingethet forteller deg at det faktum at dersom du hadde lett litt mer så kunne du kanskje funnet det nye termometeret, ikke skal ha noe å si for hvorvidt du forkaster nullhypotesen. Kun den målingen du faktisk tok skal ha noe å si for hva du tror om din kroppstemperatur.

Disse eksemplene skulle formidle innholdet i de to prinsippene. De presise definisjonene er å finne i Birnbaums artikkel og i de fleste bøker om bayesiansk statistikk, se for eksempel Robert (2007:13–18). Poenget er at disse to prinsippene er lette å akseptere, og at dersom man aksepterer dem følger det at man aksepterer likelihoodprinsippet. For mange statistikere har dette vært viktige argumenter for bayesianske statistiske metoder, fordi slike metoder ikke bryter likelihoodprinsippet.⁵

5

For å se at bayesianske metoder ikke bryter med likelihoodprinsippet må vi se litt nærmere på hva likelihoodprinsippet egentlig sier. Ovenfor skrev jeg at ifølge likelihoodprinsippet skal støtten de observerte data gir til konkurrerende hypoteser *kun avhenge av data gjennom hvor sannsynlige de observerte data er under hver av de konkurrerende hypotesene*. Hvor sannsynlige data er under en hypotese kan vi uttrykke ved den betingede sannsynligheten

$$P(\text{data}|\text{hypotese}).$$

Denne sannsynligheten leser vi som 'sannsynligheten for data gitt hypotese'. La oss skrive H for en vilkårlig hypotese (på samme måte som vi ofte lar x symbolisere et vilkårlig reelt tall). Hvis vi igjen leser den kursiverte setningen ovenfor skjønner vi at støtten data gir til H kun skal avhenge av data gjennom $P(\text{data}|H)$ betraktet som en funksjon av H. I statistikken kaller man denne funksjonen likelihoodfunksjonen og skriver $L(H) = P(\text{data}|H)$ for å understreke at den er en funksjon av hypotesene der data holdes konstant. Men dette betyr at dersom to forskjel-

lige eksperimenter gir to likelihoodfunksjoner, si $L_1(H)$ og $L_2(H)$, og disse to har den egenskap at uansett hvilken hypotese H man ser på er ratioen $L_1(H)/L_2(H)$ konstant, så skal de observerte data lede til de samme konklusjonene om våre konkurrerende hypoteser.

5.1

Bayesiansk statistikk har sitt navn fra den presbyterianske presten og matematikeren Thomas Bayes (1702–1761). Denne retningen i statistikken forbindes gjerne med det syn at sannsynlighet uttrykker et menneskes subjektive tro på at en påstand er sann. Ta ytringen ‘Med 30 prosent sannsynlighet vil singulariteten inntreffe i min levetid’. Her knyttes det en sannsynlighet til påstanden ‘singulariteten inntreffer i min levetid’, noe som krever et ikke ubetydelig tankesprang dersom man sverger til en frekventistisk forståelse av sannsynlighet. For at ytringen skal være meningsfull må man se for seg at vi lever i ett av et uendelig antall sammenlignbare universer, og at i 30 prosent av disse inntreffer singulariteten⁶.

Eller hva med en kvinne som sier ‘jeg er 60 prosent sikker på at jeg er gravid’. Enten er denne kvinnen gravid, eller så er hun det ikke, hun kan jo ikke være noe midt i mellom. Derfor, gitt at man holder seg til en strengt objektiv tolkning av sannsynlighet, er det ikke lett å si hva de 60 prosentene denne kvinnen knytter til påstanden ‘jeg er gravid’ egentlig betyr. Løsningen på disse vanskelighetene er selvfølgelig å forkaste den strengt frekventistiske tolkningen av sannsynlighet, og tillate at sannsynlighet uttrykker noe subjektivt.

La oss nå si at kvinnen som tror hun er gravid (med 60 prosent sikkerhet) tar en graviditetstest som forteller henne at hun er gravid. En graviditetstest er som kjent ikke helt sikker, så kvinnen lurer nå på hva hun bør tro om påstanden ‘jeg er gravid’ i lys av den positive testen. Bayes’ teorem forteller henne hvordan hun skal oppdatere det hun tror om verden i lys av data. La H_0 være hypotesen ‘jeg er gravid’ og H_1 dens nektelse og la D være data, som for kvinnen her er en positiv graviditetstest. Kvinnens subjektive sannsynlighet for graviditet før hun tok testen var $P(H_0) = 0.6$, sannsynligheten hun nå er interessert i er $P(H_0|D)$, sannsynligheten for H_0 gitt data. Bayes’ teorem forteller henne at den riktige måten å gå fra $P(H_0)$ til $P(H_0|D)$ er

$$P(H_0|D) = \frac{P(D|H_0)P(H_0)}{P(D|H_0)P(H_0) + P(D|H_1)P(H_1)}$$

Sannsynligheten $P(H_0)$ kalles en a priori-sannsynlighet, eller bare en prior, mens $P(H_0|D)$ kalles a posteriori-sannsynligheten, en posterior. $P(D|H_0)$ er sannsynligheten for at testen forteller kvinnen at hun er gravid gitt at hun faktisk er gravid, mens $P(D|H_1)$ er sannsynligheten for at testen forteller henne at hun er gravid gitt at hun i realiteten ikke er gravid. Nevneren er altså den totale sannsynligheten for en positiv test. I medisin kalles gjerne sannsynligheten $P(D|H_0)$ for en test sin sensitivitet, mens sannsynligheten $1 - P(D|H_1)$ er testens spesifisitet, begge tallene står gjerne på innpakningen til en graviditetstest.

Hvis vi nå spoler litt tilbake og tenker oss at kvinnen har to forskjellige graviditetstester liggende, test A og test B. På pakningen til test A står det at testen har en sensitivitet på 92 prosent og en spesifisitet på 72 prosent. Test B har sensitivitet og spesifisitet på henholdsvis 80 og 75.7 prosent. Kvinnen lurer på hvilken test hun skal bruke, men skjønner etter kort tid at siden hun bruker Bayes’ teorem har hennes valg av graviditetstest ingenting å si for den sannsynligheten hun er interessert i, nemlig posterior-sannsynligheten $P(H_0|D)$. Hvordan tenkte denne kvinnen? Prøv å forstå hvordan hun tenkte før du leser videre!

Jo, hun så at ratioen av sannsynlighetene $P_A(D|H)$ og $P_B(D|H)$ forbundet med test A og B er lik 1.15 for både H lik H_0 og H lik H_1 . Siden hun skal bruke Bayes’ teorem og $P_A(D|H) = 1.15P_B(D|H)$ for både H_0 og H_1 , ser hun at faktoren 1.15 opptrer i både teller og nevner og faller mot hverandre. Dermed er posterior-sannsynligheten den samme uansett hvilken test hun velger å bruke. Med andre ord, når hun bruker Bayes’ teorem er likelihoodprinsippet automatisk oppfylt.

5.2

For å regne ut den nye sannsynligheten $P(H_0|D)$ var den kanskje gravide kvinnen nødt til å gjøre seg opp en mening om sannsynligheten for at hun var gravid før hun tok graviditetstesten. Hun måtte anslå $P(H_0)$, for uten denne sannsynligheten ville jo høyresiden av ligningen inneholdt to ukjente.

At denne kvinnen var villig til å anslå en sannsynlighet fremstår som ganske uskyldig, men når det kommer til vitenskapen er ting annerledes. Tenk for eksempel på John-Arne Røttingen fra Folkehelseinstituttet og hans lag av forskere som dro til Guinea under ebolae epidemien der for å teste ut en ny vaksine. Mange ville nok stusset over (de gode) resultatene dersom Røttingen og hans forskerkolleger hadde gjort seg opp en mening om hvor god den nye ebolavaksinen var før avreise, for deretter å la denne formodningen om den nye ebolavaksinens effektivitet påvirke

ke resultatene de tok med seg hjem. Vitenskapens ideal om objektivitet gjør at et slikt subjektivt innslag fremstår som problematisk, så problematisk at mange, de aller fleste kanskje, heller velger å leve med at de bryter likelihoodprinsippet enn å måtte trekke subjektive a priori-sannsynligheter inn i objektive vitenskapelige analyser⁷.

Litt forenklet finnes to bayesianske svar på denne anklagen om subjektivitet. Det første anerkjenner at jo, det er et subjektivt element her, men det er det i all frekventistisk statistikk også. Forskjellen er bare at frekventistene skjuler sine subjektive antagelser, mens bayesianerne er åpne om dem. Det andre tar anklagen innover seg og forsøker å gjøre den bayesianske statistikken mer objektiv. En såkalt objektiv bayesianer trenger en a priori-sannsynlighet som reflekterer at hun ikke vet noen ting.

La oss se på myntkastene igjen. Du vil gjøre en bayesiansk slutning om kronestykkets tilbøyelighet til å vise mynt. Vi kaller denne tilbøyeligheten for θ ('theta'). Den greske bokstaven θ representerer altså et tall mellom null og én som du ønsker å si noe om ved hjelp av et eksperiment, for eksempel tolv myntkast. For å kunne bruke Bayes' teorem til å si noe om θ gitt myntkastene, trenger du en formodning om hvor mellom null og én θ ligger. Vanligvis har vi en sterk mistanke om at θ ligger nær $1/2$ et sted, men siden du nå forsøker å være objektiv vil du ikke la slike ting påvirke dine resultater. En intuitiv måte å uttrykke at man ikke vet noen ting om θ på er ved å si at alle verdier av θ er like sannsynlige. Ingen verdier er da a priori favorisert. En slik prior for θ tilsvarer at den kanskje gravide kvinnen mener at sannsynligheten for at hun er gravid er fifty-fifty, $P(H_g) = P(H_n) = 1/2$. Kaster du et blikk på Bayes teorem igjen vil du se at dette medfører at prior-sannsynlighetene i nevneren kan flyttes utenfor summen, og forsvinner mot prior-sannsynligheten i telleren. Det samme skjer i myntkasteksempelet. Vi ender altså opp med en posterior-sannsynlighet som ikke avhenger av prior-sannsynlighetene. Våre slutninger fremstår derfor som objektive og vi kan hevde at resultatene ikke er påvirket av våre subjektive oppfatninger om kronestykkets tilbøyelighet til å vise mynt.

Men vi tar feil. Et mye brukt mål i epidemiologi, medisin og gambling er odds. Odds for en hendelse er sannsynligheten for at hendelsen inntreffer delt på sannsynligheten for at hendelsen ikke inntreffer. Odds for at kronestykket viser mynt er altså $\theta/(1-\theta)$. Nå, siden vi som forsøksvis objektive bayesianere hevder at vi ikke vet noen ting om θ , er det åpenbart at vi heller ikke bør vite noe om

$\theta/(1-\theta)$. Problemet er at å si at alle verdier av θ er like sannsynlige ikke oversettes til at alle verdier av odds $\theta/(1-\theta)$ er like sannsynlige, tvert imot, noen verdier av odds er mer sannsynlige enn andre. Vi lar altså et subjektivt element snike seg inn i resultatene når vi blir bedt om å rapportere våre anslag på odds for mynt, selv om vi hevder å være helt objektive hva gjelder sannsynligheten for mynt. Dette er bare en av flere utfordringer med forsøk på objektiv bayesiansk statistikk. For en god oversikt over dette forskningsfeltet se Berger (2006).

6

Jeg håper at denne teksten har gitt leseren et visst innblikk i noen av statistikkens grunnlagsproblemer. I disse problemene møtes epistemologien, matematikken og den anvendte empiriske vitenskapen på en fascinerende måte, og en titt på litteraturen avslører at dette er et felt fullt av åpne filosofiske spørsmål. Egentlig er det ganske underlig at spørsmålet om hvilke prinsipper som skal ligge til grunn for hvordan vi bør forme våre oppfatninger om verden i lys av observasjoner ikke har fått et endelig svar, i alle fall ikke et svar de fleste kan enes om.

6.1

Om man vil lese mer anbefaler jeg å sjekke ut noen av kildene i referanselisten. Et kult sted å starte (for å se et eksempel på hva ting jeg har diskutert har å si for vitenskapen) er diskusjonen om de statistiske metodene fysikerne ved CERN brukte for å fastslå Higgs-bosonets eksistens. Den finner du her <https://bayesian.org/forums/news/3648>. Jeg ble gjort oppmerksom på denne diskusjonen i en veldig tilgjengelig og god artikkel av Sprenger (2016). Den inneholder også en introduksjon til idéene til vitenskapsfilosofen Deborah Mayo, se f.eks. Mayo og Spanos (2010). Et omfattende forsvar for likelihoodprinsippet finnes i Berger and Wolpert (1988). Tidsskriftet Rationality, Markets and Morals hadde i 2011 et spesialnummer med tittelen Statistical Science and Philosophy of Science: Where Do (Should) They Meet in 2011 and Beyond?. Bidragene er åpent tilgjengelig her <http://www.rmm-journal.de/htdocs/st01.html>. For et kortfattet historisk overblikk over uenighetene i statistikkfaget se Lillestøl (2014). En ny og meget god bok om en tredje tilnærming til statistikk (som jeg ikke har diskutert i denne teksten) er Schweder og Hjort (2016). En motivasjon for bayesiansk statistikk bygd opp rundt subjektiv sannsynlighet finnes i Savage (1972) og Bernardo og Smith (1994).

LITTERATUR

- Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer.
- . 2006. The Case For Objective Bayesian Analysis. *Bayesian Analysis*, 1(3):385–402.
- Berger, J. O. og Wolpert, R. L. 1988. The Likelihood Principle. *Lecture notes-Monograph series*, 6:iii–199.
- Bernardo, J. M. og Smith, A. F. (1994). *Bayesian Theory*. John Wiley & Sons.
- Birnbaum, A. 1962. On the Foundations of Statistical Inference. *Journal of the American Statistical Association*, 57(298):269–306.
- de Finetti, B. (2017(1975)). *Theory of Probability: A critical introductory treatment*. John Wiley & Sons.
- Deutsch, D. 1997. *The Fabric of Reality. The Science of Parallel Universes*. New York: Allen Lane.
- Efron, B. 1986. Why Isn't Everyone a Bayesian? *The American Statistician*, 40(1):1–5.
- . 1998. R.A. Fisher in the 21st Century. *Statistical Science*, 95–114.
- Hacking, I. 1965. *Logic of Statistical Inference*. Cambridge University Press.
- Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge university press.
- Lillestøl, J. 2014. Statistical inference: Paradigms and controversies in historic perspective. Technical report, Norsk Statistisk Forening.
- Lindley, D. V. 1990. The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics. *Statistical Science*, 44–65.
- Mayo, D. G. og Spanos, A. 2010. *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*. Cambridge University Press.
- Papineau, D. 2012. *Philosophical devices: Proofs, probabilities, possibilities, and sets*. Oxford University Press.
- Robert, C. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer.
- Savage, L. J. 1972. *The Foundations of Statistics*. Dover Publications.
- Schweder, T. and Hjort, N. L. 2016. *Confidence, Likelihood, Probability*. Cambridge University Press.
- Sprenger, J. 2016. Bayesianism vs. Frequentism in Statistical Inference. In Hájek, A. and Hitchcock, C., editors, *The Oxford Handbook of Probability and Philosophy*, 382–405. Oxford University Press.

NOTER

- ¹ Min oversettelse. Hacking (1965, s. 1) skriver: 'The problem of the foundation of statistics is to state a set of principles which entail the validity of all correct statistical inference, and which do not imply that any fallacious inference is valid.'
- ² Nesten, i alle fall. Se for eksempel Jaynes (2003), Savage (1972) og de Finetti (1975).
- ³ Se for eksempel Papineau (2012).
- ⁴ Som sitert i Berger (1985:29).
- ⁵ Det finnes andre metoder som ikke bryter med likelihoodprinsippet, men de bayesianske er de mest utbredte og velutviklede.
- ⁶ For en del fysikere og filosofer er kanskje ikke dette så mye å svelge? Se for eksempel Deutsch (1997).
- ⁷ For mer om denne objektiv vs. subjektiv-debatten, se for eksempel Lindley (1990) og Efron (1986, 1998).