

# Bayes' fantastiske formel

ERIK VESTERGAARD, Haderslev Katedralskole

I efteråret 2014 kom en god biologikollega til mig og foreslog et emne til en fælles elev, som gerne ville skrive SRP i matematik-biologi. Biologilæreren foreslog noget med DNA, retsgenetik og Bayes' formel. Han var nemlig kommet i besiddelse af et dokument fra Københavns Universitets hjemmesider med et forslag til et SRP-emne med netop dette tema (se [3]). Selv havde jeg ikke hørt om denne anvendelse af Bayes' formel tidligere. Jeg fandt noget materiale om Bayes' formel til eleven i vores matematikdepot. Der var tale om brugbart, omend noget "støvet" materiale fra gamle gymnasiehæfter fra 80'erne eller 70'erne og lidt fra videregående uddannelser. Det var ikke helt let og strømlinet at gå til. Og så var der dokumentet fra KU. Den matematiske del var kort og viste sig desværre at være fejlbehæftet. Derudover er den konkrete form af Bayes' formel i dokumentet ikke autentisk, skulle det senere vise sig. Idéen med at bruge Bayes' teknik var dog helt fin. Eleven fik skrevet sin opgave i december.

Jeg var nu blevet så nysgerrig angående Bayes' formel og dens anvendelser, at jeg i det nye år begyndte at studere emnet nærmere. Jeg ville gerne udrede problemerne i dokumentet fra KU og kigge på *Bayesianske netværk*, som også var omtalt i artiklen. Alle matematiklærere i gymnasiet har vel på et eller andet tidspunkt stiftet bekendtskab med Bayes' formel, måske endda i gymnasiet, hvis de hører til den ældre generation, men det er nok et hurtigt og forbigående bekendtskab for de fleste. Det var det også for mit vedkommende. Det er egentligt lidt ærgerligt, for emnet involverer en smuk teori med betingede sandsynligheder, som har vide anvendelser. Jeg har efterfølgende udarbejdet en større note, *Sandsynlighedsregning – Bayes' formel og Bayesianske netværk* om emnet med teori, eksempler, anvendelser og opgaver. Noten kan downloades fra min hjemmeside – se [2]. Interesserede er velkomne til at bruge den i undervisningen.

## Thomas Bayes

Der er store huller i den historiske viden om englænderen *Thomas Bayes* (1701? – 1761), født i London. Man ved dog, at han i 1719 blev optaget på University of Edinburgh, hvor han studerede logik og teologi. Selv om Bayes må have studeret matematik på et tidspunkt er der imidlertid ikke noget sikkert vidnesbyrd om, at det foregik på universitetet i Edinburgh. Som sin far blev han viet som præst (Nonconformist).

Thomas Bayes publicerede på intet tidspunkt artikler i matematik mens han levede, men han må have gjort sig bemærket i dette felt, for i 1742 blev han udnævnt som *Fellow of the Royal Society*. I 1763, knap tre år efter Bayes død, sendte vennen Richard Price en artikel forfattet af Bayes til Royal Society. Artiklen med titlen *An Essay towards solving a Problem in the Doctrine of Chances* blev optaget i *Philosophical Transactions of the Royal Society of London*.



Man kender ikke noget billede, som med sikkerhed forestiller Thomas Bayes. Således er det tvivlsomt, om ovenstående billede gør det. Alligevel har jeg valgt at vise det, da det ofte bliver vist i forbindelse med Bayes – første gang i en bog fra 1936.

Det paradoksale er, at den artikel, som har gjort hans navn udødeligt her mere end 250 år efter hans død, ikke efterlod noget særligt indtryk hos hans samtidige. Idéerne i hans artikel blev senere genopdaget – sandsynligvis uafhængigt af Bayes – af den store franske matematiker Pierre-Simon Laplace (1749 – 1827). Laplace videreudviklede idéerne, og emnet blev senere betegnet "invers sandsynlighed".

## Bayes' formel

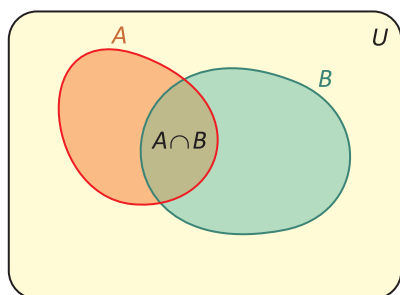
Jeg vil forlade historien og i det følgende skitsere den moderne måde at opskrive og behandle Bayes' formel på. Der vil blive lagt mere vægt på det intuitive end det formelle. Den interesserede læser kan finde en mere stringent opbygning i min note. Det skal dog nævnes, at vi for simpelhedens skyld antager at have at gøre med et endeligt sandsynlighedsfelt. Man indfører den *betingede sandsynlighed for hændelsen A givet hændelsen B*, betegnet  $P(A|B)$ , som følgende forhold:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

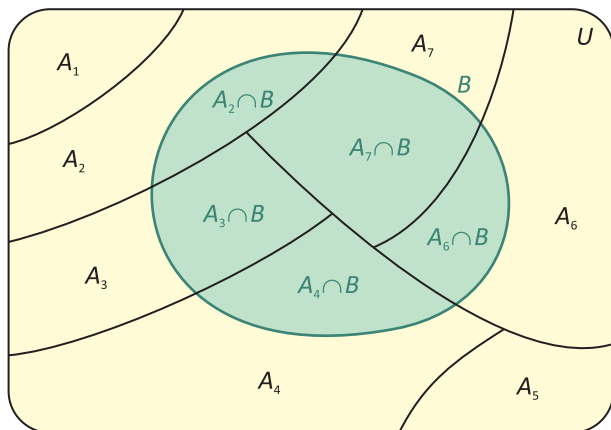
Det giver mening, for hvis man modtager *information*, at *B* er indtruffet, så "indskrænkes fokus" til mængden *B* (se figuren næste side). Sandsynligheden for at *A* også indtræffer, er derfor  $P(A \cap B)$ , og den sættes i forhold til sandsynlighe-

den for  $B$ . Ganger vi med  $P(B)$  på begge sider i definitionen, fås  $P(A \cap B) = P(A|B) \cdot P(B)$ . Venstresiden er uændret, hvis vi bytter rundt på hændelserne  $A$  og  $B$ . Derfor er højresiden det også. Vi har dermed:

$$\begin{aligned} P(A|B) \cdot P(B) &= P(B|A) \cdot P(A) \\ \Leftrightarrow P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B)} \end{aligned} \quad (2)$$



Identiteten efter ensbetydende i (2) er intet mindre end *Bayes' formel* (på simpel form). At en formel, der ligger så tæt op ad selve definitionen af betinget sandsynlighed, skal vise sig at få så enorm en betydning i udviklingen af en gren af sandsynlighedsregningen, kan virke næsten mirakuløst. Hvad formelen gør er at sammenknytte en betinget sandsynlighed med dens *inverse* betingede sandsynlighed!



Den version af Bayes' formel, som vi normalt kender, er den, hvor man har foretaget en klassedeling af udfaldsrummet  $U$  (figur), dvs. har en række indbyrdes disjunkte mængder  $A_1, A_2, \dots, A_n$ , hvis foreningsmængde er hele  $U$ . Da vil  $A_1 \cap B, A_2 \cap B, \dots, A_n \cap B$  være en klassedeling af  $B$ , hvorfor vi har:

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i) \quad (3)$$

hvor (1) er benyttet i sidste trin. Det giver følgende version af Bayes' formel:

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)} \quad (4)$$

Før vi kigger på et eksempel, skal vi præsentere en tredje udgave af Bayes' formel, nemlig *Bayes formel på odds form*. Lad  $H$  og  $E$  være to hændelser i et endeligt sandsynlighedsfelt. Da gælder:

$$\underbrace{\frac{P(H|E)}{P(H^c|E)}}_{\text{a posteriori odds}} = \underbrace{\frac{P(E|H)}{P(E|H^c)}}_{\text{Bayes' faktor}} \cdot \underbrace{\frac{P(H)}{P(H^c)}}_{\text{a priori odds}} \quad (5)$$

hvor et  $c$  påhæftet en hændelse hentyder til den komplementære hændelse, der ikke overraskende har sandsynligheden  $P(H^c) = 1 - P(H)$ . Identiteten (5) fås nemt ved at bruge (1) til at finde udtryk for henholdsvis  $P(H|E)$  og  $P(H^c|E)$  og efterfølgende dividere og reducere. Vi skal i det senere afsnit "Bayes i retsvidenskab" se en anvendelse af denne udgave af Bayes' formel.

### Test for sygdom

Der findes utallige interessante anvendelser af Bayes' formel. Jeg har valgt at præsentere et meget illustrativt eksempel, som også kan benyttes i forbindelse med faget biologi. Et screeningsprogram for en given sygdom sættes i gang for en større persongruppe, uden at der er tale om forudgående symptomer. En given person herfra får oplyst, at dennes test viser positiv. Det vides, at 1 ud af 1000 indbyggere i befolkningen har sygdommen. Som det er tilfældet for alle mulige andre testmetoder, er den pågældende ikke perfekt: I 2 % af testene fås en *falsk-positiv* og i 5 % af tilfældene en *falsk-negativ* test. Hvad er sandsynligheden for, at personen rent faktisk har sygdommen?

*Løsning:* Der er to ting i spil her: Om personen har sygdommen eller ej samt om testen viser positiv eller negativ. Vi indfører de to hændelser:

- $T$ : "Testen viser positiv"
- $S$ : "Personen har sygdommen"

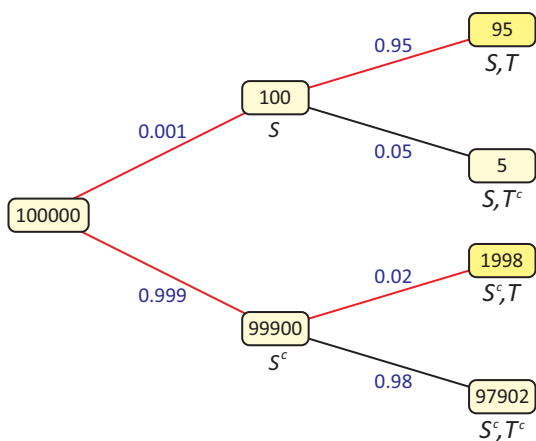
Oplysningen om falsk-positiv giver os:  $P(T|S^c) = 0,02$ , hvoraf vi har  $P(T^c|S^c) = 1 - P(T|S^c) = 1 - 0,02 = 0,98$ . De øvrige sandsynligheder overlades til læseren. Vores opgave er at bestemme sandsynligheden for, at personen har sygdommen, men med viden om at testen er positiv. Vi skal med andre ord beregne den betingede sandsynlighed  $P(S|T)$ . Det er oplagt at benytte Bayes' formel (4) i tilfældet med klasseinddelingen  $U = S \cup S^c$  af  $U$ .

$$P(S|T) = \frac{P(T|S) \cdot P(S)}{P(T|S) \cdot P(S) + P(T|S^c) \cdot P(S^c)}$$

$$= \frac{0,95 \cdot 0,001}{0,95 \cdot 0,001 + 0,02 \cdot 0,999} = 0,045$$

Resultatet er nok overraskende for de fleste. Her har personen fået at vide, at testen er positiv og at kun 1 ud af hver 1000 personer har sygdommen. Alligevel er sandsynligheden for at have sygdommen, med den nye viden om at testen er positiv, mindre end 5%! Det er godt nyt. Der skal nye undersøgelser til for at afklare, om personen rent faktisk har sygdommen!

Man kan få et indblik i, hvad der er årsagen til den lave sandsynlighed ved at tegne et hændelsestræ. Lad os sige, at vi ønsker at udregne, hvordan situationen vil se ud for en by på 100000 indbyggere, hvor alt foregår gennemsnitligt efter sandsynlighederne. Vi starter med at dele ud i to grene, alt efter om personen har sygdommen eller ej. I den forbindelse ser vi, at  $0,001 \cdot 100000 = 100$  har sygdommen, hvorimod der er  $0,999 \cdot 100000 = 99900$ , som ikke har sygdommen. Vi videreinddeler nu i grene efter om personen har en positiv eller negativ test. Her anvendes de betingede sandsynligheder. For eksempel vil der være  $0,95 \cdot 100 = 95$ , som *både* har sygdommen *og* tester positiv.



Vi bemærker, at der er ganske mange personer, som tester positive, men ikke har sygdommen. En lille procent af et stort tal, her 2% af 99900, giver i dette tilfælde et pænt stort tal. Det er årsagen til den overraskende lille chance for at have sygdommen, selv om man tester positiv. Der er  $95 + 1998 = 2093$ , der testes positive, som har sygdommen, en andel på  $95/(95 + 1998) = 0,045 = 4,5\%$ .

### Bayes i retsvidenskab

Bayes' formel har også haft sit indtog i retsvidenskab. Det er der adskillige eksempler på allerede i 90'erne. Det er ikke altid foregået uproblematisk. Faktisk har matematik mødt me-

gen modstand i dommer- og advokatkrede. Her føler man, at man mister kontrollen over sagerne, hvilket man til dels godt kan forstå. Problemet er imidlertid, at der også i argumentationen i retssalene foregår brud på de alment logiske regler, og her er det Bayes' formel, som sætter det klareste spotlight på problemet. Der findes flere typiske fejlslutninger i retssalene, hvoraf det mest kendte er *anklagerens fejlslutning (Prosecutor's Fallacy)*. Lad os for eksempel antage, at der er fundet blod på et gerningssted foruden offerets eget blod. Kun 1% af befolkningen har den pågældende blodtype. En mistænkt er anholdt, og han har netop denne blodtype.

Der er 1% chance for at tiltalte ville have samme blodtype, hvis han var uskyldig.



Der er 1% chance for at tiltalte er uskyldig.



Der er 99% sandsynlighed for at tiltalte er skyldig.



*Analyse:* Hypotesen  $H$  og viden  $E$  er følgende hændelser:

$E$ : Anklagedes blodtype matcher blodet fra gerningsstedet

$H$ : Anklagede var *ikke* på gerningsstedet

Faktisk er sidstnævnte følgeslutning i rammen korrekt, men da førstnævnte er forkert, bryder argumentationen sammen. Anklageren tror fejlagtigt, at  $P(H|E) = 0,01$ . Havde det været rigtigt, ville der rigtignok gælde dette:  $P(H^c|E) = 1 - P(H|E) = 1 - 0,01 = 0,99$ , altså en sandsynlighed på 0,99 for at anklagede var på gerningsstedet, givet viden om at blodtyperne matcher. Vi antager her, at det at have været på gerningsstedet er ækvivalent med at være skyldig! Men anklageren tager altså fejl. Det der gælder er:  $P(E|H) = 0,01$ . På engelsk går forvekslingen under betegnelsen *the fallacy of the transposed conditional* eller i denne sammenhæng: *The Prosecutors Fallacy*. Anklageren får altså vendt rundt på hændelserne i den betingede sandsynlighed. Vi kan dog godt forsøge at vurdere den "omvendte betingede sandsynlighed", nemlig ved at anvende Bayes' formel. Hertil får vi brug for a priori sandsynligheden  $P(H)$ . I fraværet af anden viden antager vi, at alle 5000 mandlige indbyggere i den lille by er mistænkte, og at de er lige sandsynlige gerningsmænd. Sidstnævnte kan naturligvis diskuteres; man kan eventuelt indskrænke til et bestemt aldersinterval. Foreløbig gør vi det dog simpelt. Sandsynligheden for at tiltalte var på gerningsstedet er dermed  $1/5000$ , og sandsynligheden for at han ikke var der, fås ved at trække første sandsynlighed fra 1:  $P(H^c) = \frac{1}{5000}$  og  $P(H) = \frac{4999}{5000}$ . Vi får af Bayes' formel (4):



$$\begin{aligned}
 P(H|E) &= \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|H^c) \cdot P(H^c)} \\
 &= \frac{0,01 \cdot \frac{4999}{5000}}{0,01 \cdot \frac{4999}{5000} + 1 \cdot \frac{1}{5000}} \\
 &= 0,980
 \end{aligned}$$

og dermed fås  $P(H^c|E) = 1 - P(H|E) = 1 - 0,980 = 0,020$ . Altså en sandsynlighed på blot 2,0 % for at anklagede er den skyldige, ikke 99 %!

Anklagerens fejlslutning er meget nærliggende og begås ikke bare af advokater, dommere og jurymedlemmer, men også af eksperter i et svagt og uopmærksomt øjeblik. Det er dermed et alvorligt problem for retssikkerheden, når beviserne overvejende er af statistisk art! I *Sally Clark-sagen* i England blev der begået flere fejl, herunder sandsynligvis anklagerens fejlslutning. Takket være den britiske advokat Marylin Stowe, som følte der var noget galt, blev sagen genoptaget og Sally Clark blev frifundet efter tre års fængselsophold. I kølvandet på sagen gennemgik den britiske rigsadvokat flere hundrede andre sager. Det førte til løsladelse af to tidligere dømte, hvis sag lignede Sally Clarks!



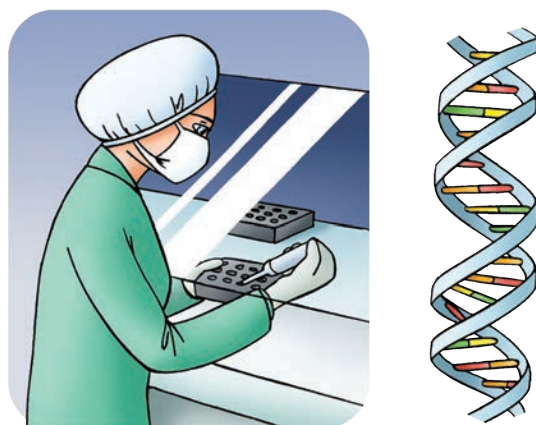
Norman Fenton, professor på Queen Mary University of London, er den ene af to forfattere til bogen [1]. Han er desuden medforfatter til diverse artikler om temaet "Bayes and the Law". [4] er blot en af disse, og den kan findes online. Ifølge Norman Fenton går dommere og advokater mere eller mindre i baglås, når de ser en formel à la Bayes' formel (4), hvorimod en del køber forklaringerne, hvis tingene forklares ved hændelsestræer såsom det, vi betragtede i forbindelse med test for sygdom. Der er i øvrigt også en tendens til, at man i det britiske retssystem foretrækker Bayes' formel på odds form (5). Argumentationen er, at man ønsker at lade jurymedlemmerne selv vurdere a priori sandsynligheden for skyld, i stedet for at overlade det til eksperter. Bayes' formel på odds form har den

egenskab, at Bayes-faktoren specifikt angiver *bevisets betydning*. Lad os for eksempel kigge på beviset med blodtype-matched tidligere i dette afsnit. Her har vi følgende Bayes' faktor:

$$\frac{P(E|H)}{P(E|H^c)} = \frac{0,01}{1-0,01} = \frac{1}{99} \quad (6)$$

*A priori odds* i (5) kan tolkes som forholdet mellem sandsynligheden for, at den anklagede er uskyldig og sandsynligheden for, at denne er skyldig – før viden om beviser. *A posteriori odds* er det samme forhold, blot opdateret med den nye viden om blodtypen! Ifølge (5) er odds for uskyld altså aftaget med en faktor 1/99, efter blodtypebeviset er afsløret. Havde man ladet  $H$  stå for skyldig, som man ofte gør, ville Bayes-faktoren have givet 99. Efter blodtypebeviset er odds for skyld altså øget med en faktor 99. Beviset har altså her en væsentlig betydning, men er ikke nødvendigvis afgørende. Bayes' faktoren kaldes også ofte for *Likelihood Ratio (LR)*.

Den britiske voldtægtssag *R v Denis John Adams* fra 1996, som er sagen behandlet i føromtalte dokument [3] fra KU, er et af de tidlige eksempler på, hvor DNA bevis alene førte til domfældelse, det til trods for at de øvrige spor i sagen talte til fordel for tiltalte. I denne sag benyttede man netop Bayes' faktorer, ikke Bayes formel direkte. En mere detaljeret behandling kan ses i min note [2].



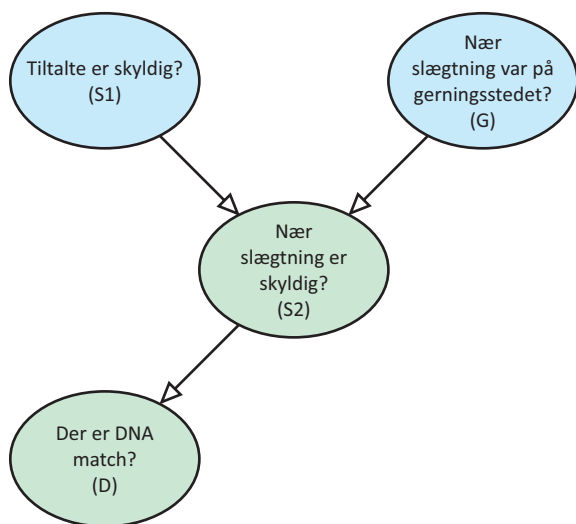
Selv om Norman Fenton ved flere lejligheder har optrådt som et ekspertvidne eller rådgiver i britiske retssager, så mener han stadig, at det britiske retssystem burde anvende Bayes mere. Det er endda hans håb, at retssystemet en gang i fremtiden rutinemæssigt vil gøre brug af de såkaldte *Bayesianske netværk*, som udvider mulighederne. Det leder os frem til næste afsnit.

### Bayesianske netværk

I begyndelsen af 1980'erne introducerede Judea Pearl, forsker i computer science og statistik fra University of California, Los Angeles, nogle specielle netværk. Formålet var, at man skulle

blive i stand til at repræsentere sandsynligheder for *tro* eller *overbevisning* (*Beliefs*) lokalt i netværket, så det tilsammen udgør et logisk sammenhængende hele. Man skulle så kunne sende ”meddelelser” igennem netværket (*belief propagation*). Pearls banebrydende arbejde blev i øvrigt i 2011 belønnet med den prestigefyldte A. M. Turing Award ”For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning”. Et *Bayesiansk netværk* (BN) er løst sagt en *orienteret acyklisk graf* med tilhørende *knudepunktstabeller* (*Node Probability Table*). I sidstnævnte står de betingede sandsynligheder, som knytter en knude sammen med dens forældrekuder. Lad os straks kigge på et eksempel.

Ved DNA-beviser er der en særlig problematik, hvis to nært beslægtede personer er mistænkte i en kriminalsag. DNA-profiler hos nære slægtninge er nemlig meget ens. I det følgende betragter vi en situation, hvor en person er tiltalt og DNA-materiale fundet på gerningsstedet matcher personens DNA. Kriminalpolitiet opdager nu, at tiltaltes broder også kan have været på gerningsstedet. Spørgsmålet er, hvilken indflydelse det har på om tiltalte er skyldig? Vores BN med knudepunktstabeller er afbildet nedenfor. I sidste tabel aflæses for eksempel, at hvis tiltalte er uskyldig og den nære slægtning var på gerningsstedet, så er der 99,9 % sandsynlighed for, at tiltalte er uskyldig. Det er oplysninger af den art, som er prop-  
 et ind i modellen ...



Tiltalte er skyldig?

False	0.9999
True	0.0001

Nær slægtning var på gerningsstedet?

False	0.9999
True	0.0001

Nær slægtning er skyldig?

Tiltalte er skyldig	False		True	
Nær slægtning var på gern...	False	True	False	True
False	1.00	0.50	1.00	1.00
True	0.00	0.50	0.00	0.00

Der er DNA match?

Tiltalte er skyldig	False		True	
Nær slægtning var på gern...	False	True	False	True
False	0.999999	0.999	0.00	0.00
True	0.000001	0.001	1.00	1.00

Det vil føre for vidt her at forklare, hvordan man regner i et sådant Bayesiansk netværk. De ret tekniske detaljer, som involverer den *simultane sandsynlighedsfordeling* (*Joint Probability Distribution*) og *marginale sandsynligheder*, kan ses i min note [2].

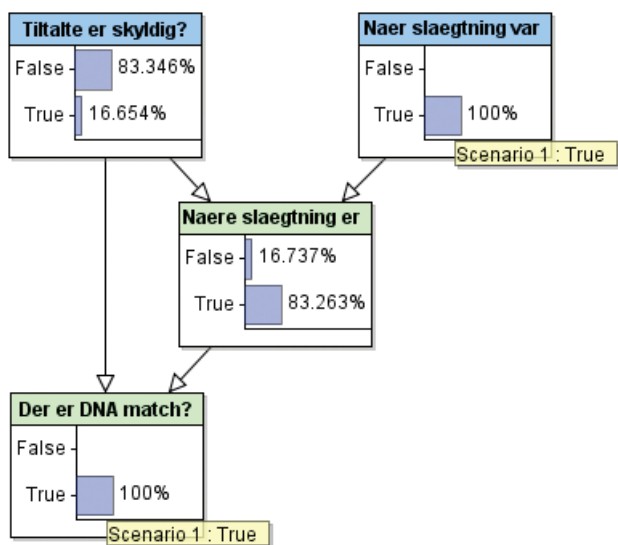
### AgenaRisk og Hugin Expert

At regne manuelt i et Bayesiansk netværk er halsløs gerning, med mindre netværket som her kun indeholder ganske få knuder. Heldigvis er der matematikere, der har udviklet algoritmer, som kan foretage beregningerne. Det er en meget mere kompleks opgave, end man skulle tro. Det har været interessant at erfare, at Aalborg Universitet her har spillet en vigtig rolle. Steffen L. Lauritzen skrev således i 1988 sammen med englænderen David J. Spiegelhalter en videnskabelig artikel, som betød en milepæl i udviklingen af algoritmer til beregning i et BN. Også Finn V. Jensen og andre fra Aalborg Universitet har leveret vigtige bidrag. I 1989 dannede en gruppe af forskere fra Aalborg Universitet firmaet Hugin Expert A/S. Firmaet er en succesfuld historie, hvor teoretisk matematik og datalogi har skabt arbejdspladser og værdi i det private erhvervsliv til glæde for Danmark. Bayesianske netværk har et bredt anvendelsesfelt, som indikeret på figuren nedenfor. På firmaets hjemmeside [L4] kan man se eksempler på, hvilke typer problemer firmaet løser for deres kunder.



AgenaRisk er et andet program til at håndtere Bayesianske netværk, udviklet af tidligere omtalte Norman Fenton, Martin Neil m. fl. Jeg har selv valgt at bruge en gratis reduceret udgave af AgenaRisk (kan downloades via [L3]), da det passer sammen med bogen [1]. Lad os gå tilbage til vores DNA-opgave. Programmet er yderst nemt at bruge. Man ”tegner” net-

værket og indtaster knudepunktstabellerne. Vi vil nu se, hvad der sker, når vi tilføjer *information* eller *evidens* til netværket, fx at der er DNA match. Vi højreklikker på knuden ”Der er DNA match” og vælger *Enter Observation > Scenario 1 > True*. Øjeblikkeligt ændrer alle de marginale sandsynligheder i knuderne sig. Sandsynligheden for at tiltalte er skyldig vokser fra a priori sandsynligheden på 0,01 % til 98,961 %. Tilføjer man yderligere den evidens, at også den nære slægtning var på gerningsstedet, så falder sandsynligheden for at tiltalte er skyldig til 16,654 %, som skærmbilledet nedenfor viser. Det er hele idéen med et BN, at man kan tilføje viden og se, hvilken betydning det har for de andre sandsynligheder. Det at bruge Bayes’ formel svarer i øvrigt til, at man har to knuder med en kant imellem og derefter tilføjer evidens til den sidste knude. Med et stort BN kan man modellere meget mere komplekse situationer. En anden fordel ved Bayesianske netværk i forhold til den mere alment kendte statistik er, at man kan repræsentere *subjektiv* viden. Ikke mere om det her.



### Pædagogisk set

I de senere år har der været stor fokus på statistik frem for sandsynlighedsregning i matematik i gymnasiet (stx). Jeg finder det til dels berettiget, at der er kommet mere statistik, idet det giver mening at give eleverne en fornemmelse for farerne ved blindt at tro på statistik. Det fornuftige i indførelsen af Chi–i–anden testen lidt inde i den nye reform kan dog diskuteres, i hvert fald læringsmæssigt. Det er formentligt et ønske fra aftagerinstitutionerne, at eleverne skal have stiftet bekendtskab med denne test. Problemet er blot, at eleverne ikke opnår nogen matematisk *indsigt* ved at lære om den. Beviserne for metodens rigtighed er alt for svære at gennemføre. Der sættes typisk bare blindt ind i en formel. De fleste elever kan ikke forklare om test–statistikken  $Q$ 's rolle eller forklare, hvad  $p$ –værdien egentligt betyder. Og hvad der er endnu værre: Stort set ingen tjekker for, at de 3 – 4 betingelser for at bru-

ge Chi–i–anden testen overhovedet er opfyldt! Testen er simpelthen for kompleks for eleverne, efter min mening. Udover den obligatoriske Chi–i–anden test, skal der på B og A niveau vælges et andet emne i statistik/sandsynlighedsregning. Her er der vel mange, som vælger normalfordelingen eller binomialfordelingen? Sidstnævnte er især velegnet, da den giver god indsigt, ligesom den er glimrende til mundtlig eksamen.

På samme måde vil jeg mene, at *betinget sandsynlighed* og *Bayes’ formel* vil være et glimrende valg:

- Emnet har en passende sværhedsgrad. Alt efter elevens/klassemens niveau kan man stoppe flere steder i teorien og stadig få et afsluttet forløb ud af det.
- Emnet involverer algebra, som er så stor en mangelvare i gymnasiet i dag. Her i form af mængdealgebra og Venn–diagrammer.
- Vil øge elevernes evne til at tænke *deduktivt*: Givet informationen, er sandsynligheden for ... Forståelse for begrebet *uafhængighed*.
- Emnet er interessant og fyldt med overraskende resultater/paradokser.
- Der er masser af anvendelser og det endda i forskellige fag.
- Velegnet til mundtlig eksamen

Derudover vil emnet naturligvis også være fint til en SRP, hvor man kan tilpasse opgavens sværhedsgrad til elevens niveau. Til de allerdygtigste kan Bayesianske netværk måske involveres. I fagkombinationen matematik–biologi vil der nok være særligt mange muligheder. Udover de aspekter, der allerede er nævnt i dette indlæg, dvs. test for sygdom og retsgenetik og DNA, vil Bayes formel kunne bruges til at besvare spørgsmål i arvelighedslære. Min note indeholder flere brugbare opgaver her.

### Litteratur

- [1] Norman Fenton, Martin Neil. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2013.
- [2] Erik Vestergaard. *Sandsynlighedsregning – Bayes’ formel og Bayesianske netværk*, 2015 (Noten kan downloades fra min hjemmeside, se [L1] nedenfor).
- [3] Flóvín Tór Nygaard Næs, Kristian Priisholm. *Retsgenetik – anvendelse af DNA–materiale i retssager*. (SRP forslag fra KU).
- [4] Norman Fenton, Martin Neil. *Avoiding Probabilistic Reasoning Fallacies in Legal Practice using Bayesian Networks*. (kan findes online)

### Links

- [L1] [matematiksider.dk/bayes.html](http://matematiksider.dk/bayes.html)
- [L2] [sites.google.com/site/bayeslegal/legal-cases-relevant-to-bayes](http://sites.google.com/site/bayeslegal/legal-cases-relevant-to-bayes) (Retssager involverende Bayes)
- [L3] [agenarisk.com](http://agenarisk.com)
- [L4] [hugin.com](http://hugin.com)