

En kort bemærkning om stokastikken

Af Jørgen Ebbesen

Ind til for nylig talte vi om statistik og sandsynlighedsregning. (Mode)ordet *stokastik* er en samlebetegnelse for de to discipliner. Stokastikken vil nedenfor blive udsat for en aldeles stedmoderlig behandling sammenlignet med den betydning, den har i tværfaglig sammenhæng i gymnasiet.

Der er sket en markant forskydning i vægtningen af de to discipliner med den seneste gymnasiereform. Før var hovedvægten på endelige sandsynlighedsfelter, gerne symmetriske og kombinatorik. Efter de seneste ændringer i læreplanen er kernestoffet i stokastikken:

simple statistiske metoder til håndtering af et datamateriale, grafisk præsentation af et statistisk materiale, empiriske statistiske deskriptorer, stikprøvers repræsentativitet og chi-i-anden test.

Og det supplerende stof skal omfatte sammenhængende et forløb *med anvendelse af yderligere mindst én type statistisk eller sandsynlighedsteoretisk model.*

Læg mærke til, at der ikke er noget krav om indførelse af statistikken på et sandsynlighedsteoretisk grundlag. Det kan man alt efter temperament begræde eller glæde sig over. Personligt var jeg glad for teorien om endelige sandsynlighedsfelter, som jeg opfatter som det eneste eksempel på anvendelse af den aksiomatisk-deduktive metode inden for gymnasiestoffet. Vi bygger flagskibet differential- og integralregning på uklare, udefinerede begreber, og det samme gælder den plane geometri. Mine elever beviser troligt hældningsformlen for en ret linje, men vi definerer aldrig en ret linje (og godt det samme, for så giver det ikke mening at udlede hældningsformlen).

Men arbejdet med de endelige sandsynlighedsfelter var tidskrævende og havde en tilbøjelighed til at lukke sig om sig selv. Vi kunne måske nok på solidt fagligt grundlag nå frem til fordelingsfunktionen for en binomial fordelt stokastisk variabel. Og bruge resultatet til at beregne sandsynligheden for, at et dige ville blive oversvømmet mindst én gang over en 20-årig periode. Og en masse andre spændende sandsynligheder. Vi nåede imidlertid aldrig frem til den statistik, som vel egentlig var grunden til, at vi indførte sandsynlighedsteorien.

Situationen er lidt den omvendte i dag (med mindre man vælger at indføre en sandsynlighedsteoretisk model). Nu er vi sikre på, at eleverne i hvert fald kommer til at arbejde med en statistisk test, men det teoretiske grundlag er plantet solidt i den blå luft. Det gør det for mig at se nærmest umuligt at sige noget fornuftigt videnskabsteoretisk om stokastikken på gymnasieniveau, al den stund at videnskaben er hældt ud med badevandet.

Skulle man være så heldig, at eleverne har kemi og/eller fysik på tilstrækkeligt højt niveau, kunne man sammenholde mekanikkens deterministiske verdensbillede med kvantemekanikkens verdensbillede og måske komme ind på Bohr/Einstein-diskussionen om tolkningen af kvantemekanikken. Og i sidste ende den overordnede realismediskussion (beskriver vi verden, som

den er, eller modellerer vi den bare efter bedste evne). Men det kræver nok et særligt engagement fra såvel lærer(e) som elever.

Der er til gengæld en masse at sige om de statistiske metoder og deres begrænsninger. Fx finder vi belæg for en påstand ved at antage, at den ikke er opfyldt, og kassere denne alternative påstand, fordi den er for usandsynlig. Det er da bemærkelsesværdigt. Og at de påstande, vi efterviser på denne måde, altid er eftervist med en vis fejlrisiko, afviger fra, hvad vi er vant til fra resten af matematikken.

Når vi udtaler os på baggrund af en stikprøve, er det en forudsætning, at den er repræsentativ for den population, vi ønsker at udtale os om. Det er et metodisk forbehold, der vil noget, al den stund, at det nærmest er umuligt at udtage en sådan stikprøve.

Man kan læse nærmere om andre metodiske spørgsmål, som det ville være naturligt at eleverne stillede sig i forbindelse med statiske undersøgelser i den form, de bliver præsenteret for dem, i skrivelsen *Om variabelbegrebet, om at prøve sig frem og om stikprøveopgaver*¹ fra fagkonsulent Bjørn Grøn, Undervisningsministeriet.

Jeg vil afslutningsvis pege på et par spørgsmål, som det vil være naturligt at tage op, hvis man anlægger en mere eksperimentel tilgang til stokastikken, nemlig i hvilken forstand, man kan sætte lighedstegn mellem frekvenser (ved mange gentagelser) og sandsynligheder. Og hvilken status simuleringer indbygget i de anvendte CAS-programmer skal tillægges.

Jeg vil personlig foretrække de klassiske modeller for kast med symmetriske mønter og terninger baseret på symmetriske sandsynlighedsfelter frem for modeller baseret på frekvenser ved gentagne kast. Skulle frekvenserne afvige fra $\frac{1}{2}$ og $\frac{1}{6}$, vil jeg være tilbøjelig til at kassere mønten og terningen/tilskrive resultaterne statistiske udsving. Jeg kunne ikke drømme om at skifte de klassiske modeller ud. De er jo ikke kun praktiske ad hoc modeller, men idealiseringer.

Og simuleringerne er bekvemme, men ikke helt uproblematisk. Tilfældige tal spiller en central rolle i stokastikken; men i hvilken forstand er de CAS-genererede tal tilfældige? Det er de selvfølgelig ikke, for der er med garanti en eller anden algoritme indbygget i softwaren, men i nogen sammenhænge er det godt nok. Det var det bare ikke dengang, jeg brugte min lommeregner tilfældige tal til at afgøre, hvem af eleverne der skulle til tavlen: eleverne valgte et tal hver, og så trykkede jeg på tilfældigt tal, indtil eleverne sagde stop. Jeg måtte opgive metoden, fordi eleverne opdagede, at lommeregneren gav de samme tal i samme rækkefølge hver gang.

¹ Kan downloades fra <http://www.emu.dk/gym/fag/ma/undervisningsministeriet/nyt/variabelbegrebet-stikproever.pdf>