Tue Tjur: Hvad er tilfældighed?

16.–19. september 1999 afholdtes i netværkets regi en konference på RUC om sandsynlighedsregningens filosofi og historie. Som ikke specielt historisk interesseret, men nok interesseret på et lidt mere teknisk niveau i sandsynlighedsregningens og statistikkens grundlag, deltog jeg i denne konference, i håb om at der ville dukke noget interessant op.

Stig Andur Pedersen har nu bedt mig om at skrive om mine indtryk. Det har jeg valgt at tage helt bogstavligt, idet jeg vil skrive om det jeg kan huske, uden at prøve at rekonstruere det jeg har glemt. Jeg vil faktisk kun skrive om en enkelt ting, en erkendelse, der for mit vedkommende gjorde konferencen til en absolut succes. Den har noget at gøre med to foredrag af Volodya Vovk og Glenn Shafer. Det var ikke noget der blev stærkt fremhævet ved disse foredrag, snarere noget der kom frem i et par sidebemærkninger. Men det satte nogle tanker igang hos mig, og det ved jeg at det også gjorde hos andre, blandt andet Teddy Seidenfeld og Flemming Topsøe.

Sandsynlighedsregningen har et grundlagsproblem, som er kendt af enhver der har forsøgt at undervise i sandsynlighedsregning. Problemet er hvad man overhovedet mener med sandsynlighed, eller med at noget er tilfældigt. Når man for eksempel siger "sandsynligheden for at få krone i et kast med en mønt er 1/2", hvad *mener* man så? I den formelle sandsynlighedsregning mener man ikke noget som helst. Der antager man det bare, dvs. man beskriver møntkast ved en model hvor de to mulige udfald har sandsynlighed 1/2, og så antager man yderligere, at hvis mønten kastes flere gange er udfaldene uafhængige. Men det er klart at man snyder både sig selv og dem man underviser, hvis man ikke siger lidt mere, for sandsynlighed er helt afgjort noget der er lige så konkret som længde, hastighed og masse. Derfor plejer man gerne (udenfor pensum, med småt osv.) at sige noget i retning af følgende:

Hvis man kaster en mønt 100 gange kan man være næsten sikker på at få mellem 33 og 78 gange krone. Hvis man kaster den 10000 gange kan man være næsten sikker på at få mellem 4721 og 5280 gange krone. Hvis man kaster den et astronomisk antal gange kan man være endnu mere sikker på at få en relativ hyppighed af kroner som ligger tæt på 1/2. Så det er, hvad sandsynligheden for en hændelse betyder: Det er den relative hyppighed af hændelsen, når man gentager det eksperiment, hændelsen er knyttet til, et meget stort antal gange.

Problemet med denne definition (som er grunden til at den altid gives udenfor referat) er, at den bider sig selv i halen. For hvis man går den på klingen, hvad betyder det så at man kan være "næsten sikker" på noget? Det betyder, hvis man vil være pedantisk, at dette "noget" indtræffer med en sandsynlighed som er næsten 1. Og så er man jo ligesom ikke kommet videre. Desuden er selve det faktum, at relative hyppigheder konvergerer, måske ikke i så høj grad en empirisk iagttagelse som en konsekvens af det formelle sandsynlighedsteoretiske resultat, der kaldes de store tals lov.

Mange — som de første vel von Neumann og Kolmogorov, og som den sidste måske Volodya Vovk — har interesseret sig for, om det ikke er muligt at give en mere håndfast forklaring. Og det som dæmrede for mig på mødet (uden at jeg tør sige, om det lige er Vovk der har opdaget det som den første) var, at det er det faktisk, i en meget mere konkret forstand end jeg troede.

Et af de anvendelsesområder som intuitionen omkring sandsynlighed kan hænges op på, er det økonomiske eller spilteoretiske. Hvad er du for eksempel villig til at betale mig for at kaste 100 mønter på gulvet og lade dig få alle dem som viser krone? Svaret er naturligvis 50 kr. Ganske vist med en masse forbehold om din risikoaversion, din manglende lyst til overhovedet at ligge på gulvet og samle penge op og derefter slæbe rundt på et bjerg af mønter osv. Men alligevel, værdien af dette tilbud er, i en eller anden fundamental forstand, 50 kr. Hvis du kunne slippe med at betale 40 kr. ville du formodentlig sige ja, hvis du skulle betale 60 ville du helt afgjort sige nej.

Man kan endda argumentere for, at denne påstand ikke nødvendigvis afhænger af sandsynlighedsteoretiske overvejelser, men alene er en konsekvens af mønters symmetri. Forestil dig, at du kommer med 50 enkroner og jeg kommer med 50 enkroner. Vi smider dem alle sammen på gulvet, du tager dem der viser krone og jeg tager dem der viser plat. Hvis det ikke er et retfærdigt spil, hvad er så retfærdighed?

Et lignende, mere generelt eksempel, er følgende. Antag at du bliver tilbudt at spille et spil, som går ud på så mange gange du har lyst at kaste en mønt. Før hvert møntkast fastsætter du din indsats. Denne indsats får du udbetalt hvis mønten viser krone, men du må til gengæld af med det samme beløb hvis mønten viser plat.

Ifølge samme intuition som ligger bag ved eksemplet med de 100 enkroner er dette et værdiløst spil. Man kan roligt vælge at spille det, risikoen for tab opvejes af chancen for gevinst. Man kan også roligt vælge at lade være, chancerne for at blive millionær på den måde er næppe store. Igen må man selvfølgelig tage nogle forbehold mod risikoaversion osv. Men disse forbehold er stort set væk hvis vi forudsætter, at du kun spiller for de småpenge du har i lommen. Et sådant krav om en begrænset formue er i øvrigt også nødvendigt, fordi man ellers kan spille med en klassisk "fordoblingsregel" for indsatsen, som helt sikkert giver positiv gevinst på et eller andet tidspunkt (nemlig første gang mønten viser krone), men som til gengæld kan indebære vilkårligt store tab undervejs.

Formelt kan situationen præciseres således: Lad $x_1, x_2, \dots \in \{0, 1\}$

betegne udfaldene af møntkastene, idet krone kodes som 1 og plat som 0. Du starter med en formue på F_0 kr. Din formue efter n'te spil kalder vi F_n. Før spil nr. n (altså før x_n er observeret) fastsætter du din indsats I_n. Det er et krav, at du skal kunne indfri din gæld hvis du taber, altså at $I_n \leq F_{n-1}$. Specielt, hvis du på et eller andet tidspunkt satser hele din formue ($I_n = F_{n-1}$) og taber, så kan du ikke spille længere. Vi vil endvidere tillade "negative indsatser", svarende til at du satser på plat i stedet for krone. I så fald kræves naturligvis at $-I_n \leq F_{n-1}$.

Åbenbart er

$$F_n = F_{n-1} + (2x_n - 1)I_n.$$

Under den sædvanlige sandsynlighedsteoretiske model for møntkast er følgen (F_n) en *martingal*, om hvilken man kan bevise adskillige ting, blandt andet at den forventede værdi af din formue efter n spil — uanset hvilken strategi du følger i dit valg af indsatser — altid vil være lig med det beløb du startede med. Men det er overhovedet ikke sådan vi tænker på det i øjeblikket. Den følgende sætning har formelt intet med sandsynlighedsregning at gøre.

Sætning. For ethvert beløb B findes en strategi (dvs. en algoritme til valg af indsats I_n, når hele forløbet til og med spil nr. $n-1$ er givet) som har følgende egenskab. Hvis følgen

$$\bar{x}_n = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}S_n$$

ikke konvergerer mod 1/2, så vil din formue F_n på et eller andet tidspunkt overstige B.

Dette resultat er, for mig at se, et meget overbevisende argument for den formelle sandsynlighedsregnings brugbarhed, i det mindste når det gælder modellering af møntkast. Det sætningen siger er sådan set, at hvis de store tals lov *ikke* gjaldt ville man gradvist kunne arbejde sig op til millionærstatus, bare ved at gå rundt og spille et banalt lodtrækningsspil med folk. Bemærk, at der ikke er tale om et af de sædvanlige tricks med at satse astronomiske ikke–eksisterende beløb. Vi satser aldrig mere end vi har. Faktisk ser man let af beviset, at sætningen holder under det yderligere krav, at alle indsatser skal holdes under en given øvre grænse. Den eneste antagelse der kan forekomme lidt urealistisk er at pengebeløb forudsættes "uendeligt delbare". I praksis ville man måske mene, at beløb skal være delelige med 1 øre (eller måske 25 øre, hvis vi taler om kontanter). Men hvis man stiller den slags krav fungerer beviset ikke.

Beviset vil vi nøjes med at skitsere, omend på en sådan måde at man uden stort besvær kan opskrive et fuldstændigt bevis på basis af skitsen. Vi begynder beskedent med at vise at det er muligt at finde en strategi, der vil få vores formue til at vokse ud over en hvilken som helst grænse, under antagelse af at den relative hyppighed $\bar{x}_n$ af kroneudfald når op over 0.51 for vilkårligt store værdier af n. Til dette formål sætter vi

$$I_n = \Delta F_{n-1}$$

for et passende $\Delta \in]0,1[$. Dvs. vi satser hver gang en fast andel Δ af vores formue. Vi får så

$$F_n = \begin{cases} (1+\Delta)F_{n-1} \text{ for } x_n = 1 \\ (1-\Delta)F_{n-1} \text{ for } x_n = 0 \end{cases}$$

og dermed

$$F_n = (1+\Delta)^{x_n}(1-\Delta)^{1-x_n}F_{n-1} = \cdots = (1+\Delta)^{S_n}(1-\Delta)^{n-S_n}F_0.$$

Antag nu at vi efter n spil har $\bar{x}_n \geq 0.51$. Så er

$$F_n = (1+\Delta)^{S_n}(1-\Delta)^{n-S_n}F_0 = \left((1+\Delta)^{\bar{x}_n}(1-\Delta)^{1-\bar{x}_n}\right)^n F_0$$

$$\geq \left((1+\Delta)^{0.51}(1-\Delta)^{0.49}\right)^n F_0 = \left((1+\Delta)^{0.02}(1-\Delta^2)^{0.49}\right)^n F_0.$$

Men som man let kan overbevise sig om (f.eks. ved at tage logaritmen og rækkeudvikle) kan man ved at vælge Δ tilstrækkeligt lille opnå at

$$(1+\Delta)^{0.02}(1-\Delta^2)^{0.49} > 1.$$

Da situationen $\bar{x}_n \geq 0.51$ kan optræde for vilkårligt store værdier af n følger det umiddelbart, at følgen (F_n) vil antage vilkårligt store værdier. Sætningens krav om en fast "målværdi" B er ikke engang nødvendig, den er kun medtaget for at gøre sætningen mere konkret.

Hermed er vi naturligvis ikke færdige. Vi skal for eksempel også sikre os, at vi når vilkårligt højt hvis $\bar{x}_n$ holder sig under $1/2$, men vilkårligt ofte kommer under (for eksempel) 0.49. Det kan vi imidlertid opnå på simpel vis ved på forhånd at dele vores formue i to dele, som hele vejen igennem holdes skarpt adskilte, ganske som om de tilhørte to forskellige spillere. Den ene del bruger vi til at satse på krone efter ovennævnte strategi, den anden del bruger vi til satsning på plat efter samme strategi. Hermed sikrer vi os, at mindst en af de to dele på et eller andet tidspunkt vokser sig så stor vi måtte ønske, hvis blot følgen $(\bar{x}_n)$ uendeligt ofte kommer uden for intervallet $[0.49, 0.51]$.

Endelig må vi selvfølgelig modificere strategien så den fungerer når bare *et eller andet* lille interval $[0.5-\varepsilon, 0.5+\varepsilon]$ forlades vilkårligt ofte af følgen $(\bar{x}_n)$. Hertil kan benyttes en lignende opdeling af den samlede formue, denne gang blot i en hel følge af positive andele $F_n = F_{n1} + F_{n2} + \ldots$, hvor andelen F_{ni} benyttes til satsning på at følgen vilkårligt ofte kommer udenfor (for eksempel) intervallet $[0.5 - 10^{-i}, 0.5 + 10^{-i}]$. Den strategi vi har konstruet i specialtilfældet $\varepsilon = 0.01$ kan naturligvis kopieres (blot med et andet valg af Δ) for en hvilken som helst værdi af $\varepsilon > 0$.

Tue Tjur
Handelshøjskolen i København
Statistikgruppen
Solhøj Plads 3
DK-2000 Frederiksberg
email tuetjur@cbs.dk