Viden Om – Vind oftere, stop i tide

Spørgsmål og svar

Jørgen Hoffmann-Jørgensen

Indhold

Risici og relevans

Steffen Andersen

Jeg så med forundring udsendelsen »Viden OM« i går på DR2. Udsendelsen handlede om optimale stopregler, sandsynligheder og risici i det hele taget, både i spil og i andre a livets facetter hvor individer tager beslutninger.

Jeg blevet meget forvirret, og overvejede om DR (+ måske den måde spørgsmålet og emnet er blevet fremlagt for eksperterne) fuldstændig har misforstået begrebet risici og forventet gevinst/udfald. Pointen med at tegne en livsforsikring, eller ikke at gå efter det »maximale« stop tidspunkt i Deal or No deal er NETOP at man ikke kan antage en gennemsnits-betragtning.

Dette begreb »Gennemsnitlig kan det ikke betale sig« som blev benyttet i flæng er fuldstændigt irrelevant. TV2 lader dig ikke spille »Deal or No Deal« 117 gange for at opnå en gennemsnitlig maksimal pris. Gud lader dig heller ikke dø 117 gange for at finde ud af om det gennemsnitlig betale sig at tegne livsforsikringen.

Hele eksemplet med den optimale stopregel på parkeringsplads er derfor også lidt sjovt. Hvad hvis min kone ligger og er i gang med at føde på hospitalet? Tør jeg så benytte den optimale stopregel? Det kunne jo være der kun var 16 ud af de optimale 17 parkeringspladser... Det kunne jo være jeg kunne komme tættere på hospitalet og nå fødslen, eller det kunne være jeg bare smed den foran porten, og risikerede at få bilen fjernet... Det er jo ikke hver dag at min kone føder, og her ville jeg nok vælge IKKE at antage en gennemsnitsbetragtning... Det kunne også være min svigermor der ventede ved biografen... ville jeg så tage en ekstra chance?

Dette er noget man lære på 1 år på ethvert økonomikursus, vist i bjerg af videnskabelige artikler og noget som de fleste med en basal intuition kan sige sig selv. Mennesker i IKKE neutrale over risici, og dette har INTET med at fejlvurdere sandsynligheder at gøre. Desuden er det stort set kun spil som Kasino og andre meget simple spil man kun spiller med sig selv hvor »rules and probabilities of the game« er kendt (selv i et spil så simpelt som Lotto kender du ikke udbetalingerne, da de afhænger af indbetalingerne). I de fleste andre af livet facetter er det hverken mulige at karakterisere udfald eller sandsynligheder, og derfor har man stort set kun sin INTUITION at handle ud fra.

Min spørgsmål er så:

> *Hvor relevant er en gennemsnitsbetragtning hvor man antager at folk forholder sig neutralt over for risici overhovedet?*

Svar

Dine spørgsmål berører nogle meget fundamentale problemer, nemlig: »Hvad er en sandsynlighed?«. Der er (mindst) tre mulige måder at fortolke sandsynligheder på:

(1) *Den frekventielle fortolkning.* Sandsynligheden for en hændelse er lig hyppigheden af hændelsen i en (lang) række uafhængige gentagelser af hændelsen. F.eks.

hvis 100 kast med en mønt giver »krone« 36 gange og »plat« 64 gange, så er den (frekventielle) sandsynlighed for »krone« lig $\frac{36}{100}$. I denne fortolkning er sandsynligheden for en hændelse, som ikke kan gentages ikke defineret. F.eks. hændelsen at Store Bælts Broen bliver nedsejlet i løbet af de næste 100 år, har ikke en frekventiel sandsynlighed, men faktisk bad Folketinget i sin tid et konsulentfirma om at udregne denne sandsynlighed. De kom med svaret 5%. Folketinget ønskede at bruge denne sandsynlighed til at afgøre, om der skulle iværksættes en række (dyre) sikkerhedsforanstaltninger ved Store Bælts Broen. Som bekendt blev disse sikkerhedsforanstaltninger iværksat – man anså risikoen for at være for stor til at kunne ignoreres. I 1930'erne forsøgte von Mises at give en præcis matematisk model for frekventielle sandsynligheder – uden synderlig held – modellen indeholdt en række uløselige logiske problemer.

(2) *Den subjektive fortolkning.* Sandsynligheden for en hændelse er lige den grad af overbevisning, som den person der fremsætter den, har om hvorvidt hændelsen indtræffer. F.eks. konsulentfirmaets sandsynlighed på 5% for en nedsejling af Store Bælts Broen i løbet af de næste 100 år, udtrykker firmaets grad af overbevisning om hvorvidt denne hændelse vil finde sted. Subjektive sandsynligheder giver mening for enhver hændelse (også selvom den ikke kan gentages), men den er subjektiv: Forskellige personer kan have forskellige subjektive sandsynligheder for samme hændelse. Nogle sandsynligheder er både frekventielle og subjektive: Ifølge mit forsikringsselskab er sandsynligheden for, at mit hus nedbrænder i løbet af næste år lig $\frac{1}{20.000}$. Forsikringsselskabets sandsynlighed er frekventiel (den er baseret på en lang række observationer af parcelhuse, der »ligner« mit hus). For mig er sandsynligheden $\frac{1}{20.000}$ subjektiv. Hændelsen »mit hus nedbrænder i løbet af det næste år« kan ikke gentages. Til trods herfor tror jeg på sandsynligheden og har tegnet en brandforsikring (det ville jeg også gøre, selvom den ikke var lovpligtig). I 1933 lavede Kolmogorov en præcis matematisk model for sandsynlighedsregningen, som i dag er næsten enerådende i sandsynlighedsregningen. Kolmogorovs model er en model for subjektive sandsynligheder: Den specificerer *ikke* hvad sandsynligheden for en hændelse er, men den giver de fundamentale regler (aksiomer) for regning med sandsynligheder. De subjektive sandsynligheder indeholder de frekventielle sandsynligheder i den forstand, at vi i Kolmogorovs model faktisk kan *bevise* at den frekventielle sandsynlighed er lig den subjektive sandsynlighed. Mere præcist kan vi bevise *de store tals lov*:

- Hvis $X_1, X_2, \ldots$ er uafhængige tilfældige tal med samme middelværdi og samme varians, da vil den »frekventielle middelværdi«; dvs. gennemsnittet $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$, være ca. lig den fælles middelværdi, når blot antallet af observationer n er tilstrækkelig stor – eller på »matematisk«:
- $P\big(\lim_{n\to\infty} \bar{S}_n = \mu\big) = 1$ hvor μ er den fælles middelværdi.

(3) *A priori fortolkningen.* A priori sandsynligheden for en hændelse er den sandsynlighed, som alle fornuftige mennesker kan enes om. F.eks. i almindelighed går man

udfra at sandsynligheden for en 6'er it et slag med en terning er lig $\frac{1}{6}$, også uden at lave forsøg med terningkast. A priori sandsynligheder er som regel baseret på ideén om »lige mulige« udfald, og begrænset til de hændelser, som alle fornuftige mennesker kan enes om. Den oprindelige model for sandsynlighedsregningen (B. Pascal & P. Fermat, 1654), er netop en model for a priori sandsynligheder. I dag er a priori sandsynligheder grundlaget for Bayes' statistik, hvor man starter med en apriori sandsynlighed, udfører et forsøg, og beregner en posteori sandsynlighed (ved hjælp af *Bayes' formel* eller varianter af denne).

For at komme til dine konkrete spørgsmål:

I »Deal no-deal« er de relevante sandsynligheder frekventielle for TV2 og den »gennemsnitlige gevinst« er en reel størrelse for TV2 – de spiller mange gange. For den enkelte deltager er »knald eller fald« – hun/han spiller kun én gang og de relevante sandsynligheder er således subjektive for den enkelte deltager – men til trods herfor må det mest rationelle være at indrette strategien således, at man maksimerer sin gennemsnitlige gevinst. I hvert fald efter min overbevisning, som selvfølgelig kan diskuteres, men jeg kan ikke se nogen anden rational strategi. Faktisk opfattede Jakob Bernoulli(1654–1705) i sin berømte bog *Ars conjectandi* (som faktisk indeholder det første bevis for *store tals lov*, og som blev udgivet i 1713), sandsynlighedesregningen som en model for »den sunde fornuft«, nemlig som en måde til at tage rationale beslutninger i en tilfælgig Verden.

Parkeringsproblemet: Som altid i sandsynlighedsteoretiske modeller er der en række underliggende antagelser. I parkeringsproblemet lyder de således: (a) Hændelserne »en P-plads er optaget« er uafhængige og har samme sandsynlighed, som er kendt, lad os sige p. (b) Antallet af P-pladser er kendt, lad os sige d. (c) Positionen af biografen (eller hospitalet) er kendt, lad os sige udfor P-plads nr. n. (d) Der er opgivet en omkostning, lad os sige A, ved ikke at komme i biografen (eller på hospitalet). Den optimale P-strategi giver et bestem tal, lad os sige r, som kan beregnes udfra størrelserne p, d, n, og A, og den optimale strategi siger: »passér de første $r-1$ P-pladser, tag dernæst den første ledige P-plads derefter«. Den præcise formel for udregningen af den kritiske værdi r er eksotisk:

$$r = 1 \wedge \operatorname{ceil}\left(\frac{\log\{(A-d-n)p^d(1-p)+p^n(1-p)-p^d\}}{\log p}\right)$$

men hvis omkostningen A ved ikke at komme i biografen (eller på hospitalet) er meget stor, så er $r = 1$, og du skal vælge første ledige P-plads.

For hændelser i den »virkelige Verden« må man evaluere sine sandsynligheder:

(1) Enten som en frekventiel sandsynlighed. Dette gøres ved observationer af en række »lignende« hændelser; f.eks. sandsynligheden for »mit hus nedbrænder i løbet af det næste år«;

(2) Eller som en subjektiv sandsynlighed. Dette kan f.eks. gøres ved simulationer på en computer, det var den måde, som konsulentfirmaet udregnede sandsynligheden for hændelsen: »Store Bælts Broen bliver nedsejlet i løbet af de næste 100 år«.

Intuitionen er ingen hjælp her, og den giver alt for ofte et urealistisk bud på sandsynligheden. Lad mig give et simpelt eksempel. En mønt kastes igen og igen, dette giver os en række af K'er (»krone«) og P'er (»plat«). Lad os betragte første forekomst af PK i den rækkefølge. Da sandsynligheden for PK er lig $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, vil man intuitivt forvente at PK kommer cirka hver fjerde gang, og at man således i »snit« skal vente 4 gange på første forekomst af PK. Dette er faktisk korrekt, men udregningen er meget lang og »tricket«. Så vidt så godt intuitionen er korrekt. Lad os betragte første forekomst af KK. Da sandsynligheden for KK også er lig $\frac{1}{4}$, er det klart at man ligeledes i »snit« skal vente 4 gange på første forekomst af KK – eller er det nu også det? Gennemføres beregningerne, som igen er lange og »tricket«, får man at den gennemsnitlige ventetid på KK er lig 6. Det er ikke så godt – intuitionen er forkert. Min intuition siger også at den gennemsnitlige ventetid på KK skal være 4, men mine beregninger siger at den er lig 6, og desuden viser forsøg med møntkast at 6 er korrekt. I sandsynlighedsregningen er der et hav af lignende eksempler, som viser at vores intuitive forestillinger om sandsynligheder er håbløst forkert og strider mod beregninger (og forsøg!). Husk også at sandsynlighedsbegrebet er et nyt begreb, som kun har eksisteret i ca. 450 år. Før den tid eksisterede begrebet overhovedet ikke, til trods for at spil har været kendt lige siden mennesket kravlede ned fra træerene og begyndte at gå på to ben. Denne lange erfaringsrække udviklede ikke en intuition om sandsynligheder.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Sandsynligheder

Per Hedegård

I udsendelsen siger Hoffmann-Jørgensen, dels at man ikke skal stole på sin intuition men på beregningerne, og at hvis sandsynligheden er $\frac{1}{6}$ for at slå en sekser i det første slag, så er den også $\frac{1}{6}$ hvis man efter 100 slag ikke har slået en sekser. Min intuition siger mig, at det sidste absolut ikke kan passe. Sandsynligheden for at slå 100 ikke-seksere i træk, givet at sandsynligheden for en sekser i hvert slag er $\frac{1}{6}$, er $\frac{1}{83 \text{millioner}}$ altså en tiendedel af sandsynligheden for at få 7 rigtige i lotto! Jeg ved også matematikernes »ærlige terning« ikke findes i virkeligheden, men er en teoretisk fiktion. Enhver rigtig terning er en fysisk virkelig ting, og det skulle være meget mærkeligt, om den var »ærlig«. Hvis jeg efter 100 slag ikke havde set en sekser, ville jeg bestemt komme i tvivl herom og sige, at sandsynligheden for at terninger var »ærlig« er ganske lille og derfor og ændre min sandsynlighed for at slå en sekser i slag nummer 101. Intuition og common sense synes i dette tilfælde at være mere korrekt end beregninger baseret på matematiske fiktioner som »ærlige terninger«.

Svar

I Lotto er en hver given kombination overordentlig usandsynlig, også den kombination der bliver udtrukket. Så usandsynlige hændelser indtræffer! Med hensyn til terningen. Dette er et gammelt problem, som går tilbage til midtena f 1700-tallet, hvor de to matematikere d'Alember og Daniel Bernoulli diskuterede om hvorvidt terningen har »hukommelse« eller ej. Daniel Bernoulli påstod, at terningen har ingen »hukommelse«, og d'Alembert påstod, at det havde den. De diskuterede problemet i mere end 30 år uden at nå til enighed. Jeg vil sige at hvis en terning ikke har vist en 6'er i 100 slag, så er den måske skæv, men på den anden side er der blevet spillet så meget terninger i løbet af de sidste cirka 6000 år (så gammel er terningen), at det ikke er usandsynligt at en »ærlig terning« på et eller andet tidspunkt i disse mange år ikke har vist en 6'er i 100 successive slag.

Med hensyn til matematiske »fiktioner«. Det er korrekt, at når vi anvender matematik (herunder sandsynlighedsregning) på den »virkelige Verden«, så bygger vi på en idealiseret fiktion. Problemet er så om fiktionen fungerer i praksis. Ved at anvende matematik, kan vi forudsige hvor og hvornår den næste solformørkelse indtræffer – og det gør den! Matematik er *ikke* en naturvidenskabelig disciplin. Matematikkens objekt er *ikke* den »virkelige Verden«, men dens objekt er vores egen tanke. Dette gør matematik til en filosofisk disciplin, som af historiske årsager i dag er henlagt til de naturvidenskabelige fakulteter. Det er noget af et under at matematikken faktisk fungerer i den »virkelige verden«, og nogle filosoffer går så langt, at de siger, at matematikken er »Naturens sprog«. Matematikken fungerer upåklageligt i naturvidenskaberne, i den forstand at den giver korrekte forudsigelser. Forsikringsselskaberne er det bedste bevis for, at sandsynlighedsregningen fungerer i den »virkerlige Verden«. Forsikring og sandsynlighedsregning opstod cirka samtidigt (i midten af 1600-tallet), og forsikring var den første seriøse anvendelse af sandsynlighedsregningen. Gå ud og se forsikringsselskabernes paladser, hvis sandsynlighedsregningens »fiktion« ikke fungerede i den »virkelige Verden«, ville de ligge i ruiner!

Med hensyn til »common sense«: Faktisk opfattede Jakob Bernoulli (1654–1705) i sin berømte bog *Ars conjectandi* (*Kunsten at gætte* udgivet i 1713), sandsynlighedsregningen som en model for »den sunde fornuft«, nemlig som en måde til at tage rationale beslutninger i en tilfældig Verden.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Spørgsmål til eksperten

Thomas Andersen

Det er nogle rigtig spændende udregninger der bliver beskrevet i programmet. Især det med terningerne.

Men jeg har nu et spørgmål rettet mod poker, som jeg selv spiller en del.

Jeg kunne godt tænke mig at vide, hvor meget held spiller ind i poker kontra skills. F.eks hvad er held procenten, hvis man spiller en Sit&Go turnering hvor blindsne starter 20/40 og stiger hvert 10. minut og der er 10 mennesker med. Og sætte det op imod et cash game, hvor blindsne ikke stiger. Er det i høj grad nemmere at vinde penge i cash games eller Sit&gos?

Svar

Poker er *ikke* et hazard-spil, i den forstand at i det lange løb vil den spiller der er bedst til at evaluere sandsynlighederne og agere korrekt udfra disse sandsynligheder vinde. Hvis to spillere er lige gode (eller lige dårlige) til at evaluere sandsynlighederne, er spillet et »fair« spil, hvilket betyder de i det lange løb begge vil ende op med en gennemsnitlig gevinst på 0 kr. Så det er umuligt at give et svar på dit spørgsmål, eller rettere svaret afhænger af hvor gode spillerne er til at evaluere de relevante sandsynligheder. Held eller uheld er fænomener, som udviskes over en lang periode. Ikke destomindre har mange spillere den erfaring, at de en given aften sidder i held eller i uheld. Dette er en korrekt observation, som skyldes den fuldstændige tilfældighed i kortfordelingen.

Lad os betragte to spillere, som spiller en række spil i løbet af en aften. I sandsynlighedsregningen siger vi en spiller sidder »i held«, hvis han/hun har ført (dvs. har haft en positiv gevinst) de meste af tiden, og sidder »i uheld«, hvis han/hun har været bagud (dvs. har haft tab) det meste af tiden. Antag, at de to spillere er lige dygtige, dvs. sandsynligheden er lig $\frac{1}{2}$ for at den ene eller den anden vinder et spil. Umiddelbart, vil man forvente, at når spillerne er lige dygtige, så vil de to spillere have ført/været bagud cirka halvdelen af spillene. I sandsynlighedsregningen kan vi bevise den såkaldte *arcus-sinus lov*, som siger at der er en stor sandsynlighed for at den ene spiller sidder »i held« og den anden sidder »i uheld«, mens sandsynligheden for at de to spillere har ført/været bagud cirka halvdelen af spillene er forbløffende lille. Mere præcist, siger *arcus-sinus loven*:

> Du spiller en række »fair« spil. Hvis F_n betegner antallet af gange, du har haft positiv gevinst i de n første spil, da vil $\frac{F_n}{n}$ være den del af tiden d har »ført« i de n første spil, og der gælder
>
> $$\lim_{n\to\infty} P\big(x \le \tfrac{F_n}{n} \le y\big) = \frac{1}{\sqrt{2\pi}} \int_x^y \frac{1}{\sqrt{t(1-t)}}\, dt \quad \forall 0 \le x \le y \le 1.$$

Hvis $\frac{F_n}{n}$ er cirka $\frac{1}{2}$, har du ført/været bagud i cirka halvdelen af spillene, hvis $\frac{F_n}{n}$ er lille, dvs. tæt på 0, har du siddet »i uheld«, hvis $\frac{F_n}{n}$ er stor, dvs. tæt på 1, har di siddet »i held«.

Arcus-sinus loven viser at den gennemsnitlige værdi af $\frac{F_n}{n}$ er lige $\frac{1}{2}$, men den viser også, at $\frac{1}{2}$ er den mest usandsynlige værdi af $\frac{F_n}{n}$, og at 0 og 1 er de mest sandsynlige værdier af $\frac{F_n}{n}$. Der betyder at den gennemsnitlige værdi $\frac{1}{2}$, *ikke* skyldes at $\frac{F_n}{n}$ er cirka $\frac{1}{2}$, men at den fremkommer som et gennemsnit af tal som er næsten lig 0 eller næsten lig 1.

Beviset for *arcus-sinus loven* er overordentlig langt og svært, men den er udtryk for at tilfældigheder opfører sig ganske anderledes end man forventer – vores intuition om sandsynligheder leder os ofte på vildspor. *Arcus-sinus loven* blev bevis omkring 1955 af en dansk sandsynlighedsteoretiker, Erik Sparre Andersen (som iøvrigt var min første lærer i sandsynlighedsregning).

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Til Jørgen Hoffmann-Jørgensen
Rasmus Østergaard Pedersen

Spiller man blackjack, roulette og lignende spil hvor oddsene er imod dig bør man vel kunne vinde. Jeg mener fx at i blackjack er sandsynligheden for at vinde omkring $\frac{1}{3}$. Spiller man nu flere spil og fortsætter med at 3 doble indsatsen vil man altid vinde mere. Altså i først spil har man sandsynligheden: $1-\left(\frac{2}{3}\right)^1 = \frac{1}{3}$, for at vinde I andet spil $1-\left(\frac{2}{3}\right)^2 = \frac{5}{9}$, altså allerede i andet spil overvejene sandsynlighed for at vinde. Fortsætter man denne tankegang lidt, vil man hurtigt se at der er så oplagt sandsynlighed for at vinde inden man kommer op på en ustyrlig indsats at man altid bør vinde.

Da spillene aldrig taber må der vel være en fejl i det jeg siger. Hvad er galt med denne tankegang, hvor man altid vil komme ud oven ud?

Svar

I blackjack er oddsene *ikke* imod dig - de er med dig. Dette skyldes at du kender »dealerens« strategi. Den optimale strategi i blackjack blev udregnet i begyndelsen af 1960'erne og den giver spillerne en gennemsnitlig gevinst på lidt under 3% af hendes/hans opsætning. Uheldigvis kræver den optimale strategi, at man kan tælle og huske mange kort (der er cirka 300 mulige kort i en »stack«), dette er dog den mindste kunst, strategien kræver at man kan udføre mange og lange beregninger på kort tid. Man skal derfor helst have en computer til rådighed og helst en såkaldt »talknuser« – en almindelig computer er for langsom.

Vedrørende dine beregninger. Der er altså sandsynlighed $\frac{1}{3}$ for at finde første, andet, tredje spil etc. Der er sandsynlighed $\frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$ for at tabe første spil og vinde andet spil. I almindelighed er der sandsynlighed $\left(\frac{2}{3}\right)^{n-1} \cdot \frac{1}{3}$ for at tabe de $n-1$ første spil og vinde det n'te spil. Disse sandsynligheder bliver overordentlig små når tallet n bliver stor. Den strategi du omtaler er en variant af den såkaldte *martinal strategi.* Det er en almindelig overtro blandt spillere, at martinal strategien er en sikker vinder strategi. Jeg vil gerne benytte lejligheden til kraftigt at advare mod strategien - i sandsynlighedsregningen kan vi vise at den med sikkerhed (dvs. 100%) fører til ruin! Se også mit svar til Martin Konstantinovitch.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

En sikker strategi ved rouletten

Martin Konstantinovitsch

Ved rouletten kan man spille på rød og sort. Gevinsten på disse to spil er indsatsen $\times 2$. I flg. min teori, så burde det altså kunne lade sig gøre at få et overskud, ved at bruge følgende metode:

§1 Sats altid på samme farve.

§2 Sats altid det dobbelte af det foregående spil.

Et eksempel: Laveste indsats er 25.

- §1. Spil laveste indsats.

  Spil 25 i første spil. Har du vundet, er gevinsten $50 - 25 = 25$, og du starter fra §1.

- §2. Har du tabt, så satser du indsats fra sidste spil $\times 2$.

  Spil 50 i andet spil. Har du vundet, er gevinsten $100 - (50 + 25) = 25$, og du starter fra §1. Har du tabt, så starter du fra §2.

- Spil 100,- i tredje spil. Har du vundet, er gevinsten $200 - (100 + 50 + 25) = 25$, og du starter fra §1. Har du tabt, så starter du fra §2.

Bordene har altid et loft, men alligevel burde sandsynligheden for at ramme loftet være mindre end sandsynligheden for at ramme rigtig farve. Har jeg fundet en skudsikker teori, eller er jeg bare blevet blind i min jagt på en nem indtjening?

Svar

Den strategi du omtaler, har været kendt i århundreder. Den går under navnet *martingalstrategien*. Ordet »martingal« har flere betydninger på dansk; f.eks.: (a) Et specielt par bukser med både livrem og seler; (b) Stormskødet til mesansejlet på et sejlskib. Mesansejlet er det agterste sejl på et tre-mastet sejlskib. Det sidder på langs af skibet, i modsætning til storsejlene, som er tværstillet. I stomvejr måtte man stryge storsejlene og lægge skibet op mod vinden, men for at roret kunne virke måtte man sætte et sejl for at give skibet egenfart – det var mesansejlet, som så blev sikret med et tovværk, som blev kaldt »martingalen«. (c) Den læderrem, som går fra en hests bringe op til hagen, og som sal forhindre hesten i at »slå« med hovedet og kaste rytteren af. (d) Den spilstrategi, som du omtaler i din email. Der er flere andre betydninger af ordet »martingal«. Fælles for dem alle, er at de er en slags sikkerhedsforanstaltning. Ordet stammer fra den franske landsby *Martingaux*, hvis indbyggere fra gammel tid er kendt som forsigtige folk.

Det er almindelig overtro blandt spillere, at martingalstrategien er en sikker vinder strategi. I sandsynlighedsregningen kan vi bevise, at martingalstrategien med sikkerhed (dvs. med sandsynlighed 100%) fører til ruin (også selv uden »loft«); medmindre man har uendelig stor kapital – og det er uendelig stor i matematisk betydning – meget stor er ikke nok! Det er vist de færreste der har uendelig stor kapital, og hvis man har kan man ikke forøge sin kapital ved at vinde: »uendelig + hvadsomhelst = uendelig«. Så lad være med at benytte martingalstrategien – den fører til sikker ruin! Som det også blev pointeret i udsendelsen, skal man betvivle sin intuition og stole på sine beregninger, men det forudsætter at man regner korrekt, og det er ikke så nemt. De beregninger der viser at martintalstrategien fører til sikker ruin er lange og komplicerede. I sandsynlighedsregningen kan vi bevise en »lov« som går under navnet *Optimal sampling* (jeg har ikke fundet en passende dansk oversættelse). Optimal sampling-loven sige (løst sagt): Hvis man har endelig kapital og spiller en række spil, hvor oddsene er imod en og man ikke er clairvoyant (en sådan række spil kaldes en *supermartingal* i sandsynlighedsegningen), så findes ingen strategi, der kan ændre odds. Den optimale strategi er simpel – men beklagelig: »Stop før du begynder«. Abraham de Moivre (1667–1754) udtrykker det meget klart i forordet til sin bog *Doctrine of Chances* (3. udgave, udgivet i 1756):

> *Your Lordship* (her hentydes til de Moivre's ven og velgører Lord Carpenter) *does easily percieve, that this doctrine is so far from encouraging Play, that it is rather a Guard against it, by setting a clear Light, the Advantages and Disadvantages of those Games wherein Chance is concerned.*

Mere præcist, sigere *optimal sampling-loven* følgende:

> Lad $X_1, X_2, \dots$ være en supermartingal af tilfældige tal, og lad σ og τ være stopstrategier, så at $\tau \le \sigma$ og σ er »optional for X_n^-«, da gælder
>
> $$E(X_\sigma \mid \mathscr{F}_\tau) \le X_\tau \ .$$

Størrelsen $E(X_\sigma \mid \mathscr{F}_\tau)$ betegner dit bedste bud på, hvad din gevinst er til tid σ baseret på al den information du har til tid τ. »Optionaliteten« af σ er et udtryk for at meget stort tab ikke er tilladt. Formlen siger så, at uanset hvordan du vælger din stopstrategi, så er det sådan, at jo længere du venter med at stoppe, jo mere taber du.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Sandsynlighed

Niels Henrik Jensen

Godt gået med parkeringspladsen.

Selv om jeg var meget god til matematik, forstod jeg aldrig rigtigt sansynlighedsberegning. Det har ærgret mig lige siden, og det har varet længe, da vi må være næsten jævnaldrende.

Kan du anbefale en bog, der kan sætte mig igang med at forstå og beregne?

Svar

Der findes ikke så meget litteratur på dansk, som er tilgængelig for lægmand uden professional hjælp, men jeg kan anbefale Erling Andersens bog *Regn på din gevinst.* På engelsk kan jeg anbefale:

(1) William Feller: *An Introduction to Probability Theory and Its Applications* udgivet af John Wiley & Sons. Bogen er fra 1950, men den kan stadig købes.

(2) Abraham de Moivre: *Doctrine of Chances.* Bogen er fra 1756, men den kan fås i dag (fotografisk optryk af originalen) på forlaget: American Mathematical Society & Chelsea. Bogen er godt nok gammel, men det er en fremragende og meget pædagogisk bog, og den når vidt omkring i sandsynlighedsregningen.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Sandsynlighed

Lars Roland Capion

Jeg har et spørgsmål til Jørgen Hoffmann-Jørgensen vedr. sandsynlighed.

Hvis der er to mulige udfald af et kast. F.eks. ved kast med en mønt, hvor der er lige stor sandsynlighed for plat som for krone, er det klart, at ligegyldig hvor mange forudgående ens udfald, der har været, er udfaldet af næste kast stadig lige tilfældigt og med lige stor chance for plat som for krone.

Men samtidig er mønten nødt til at stræbe mod en ligelig fordeling ml. plat og krone, hvis den skal være en normal mønt. Så hvis der har været et stort antal udfald af f.eks. plat, må normalfordelings-»kraften« vel på et eller andet tidspunkt begynde og »trække« mønten mod det modsatte udfald, så en ligelig fordeling af udfaldende opnåes.

Men da kastet med mønten skal være tilfældigt, er dette en umulighed, og derfor synes jeg der er en mystisk modsætning ml. møntens tilfældige udfald og den ligelige sandsynlighedsfordeling.

Kan du/I løse mysteriet? På forhånd tak!

Svar

Dit første spørgsmål: Hvis mønten har sanssynlighed $\frac{1}{2}$ for at vise »krone« og »plat«, så gælder, at ligegyldigt hvad mønten har vist de sidste 10.000 gange, da er sandsynligheden stadig lige $\frac{1}{2}$ for at vide »krone« og »plat« i kast nummer 10.001. Det er en almindelig overtro, at når blot man har tilstrækkelig mange observationer, så er alting normal fordelt. I sandsynlighedsregningen kan vi bevise den såkaldte *centrale grænseværdi lov* som løst sagt, siger at summen af mange »små« og uafhængige tilføldige tal, er approksimativt normal fordelt; dvs.: Hvis $S_n = X_{1n} + \cdots + X_{nn}$ er en sum af n »små« tilfældige tal $X_{1n}, X_{2n}, \ldots, X_{nn}$, så gælder, under en række specifikke forudsætninger, at vi har:

$$\lim_{n\to\infty} P(x \le S_n \le y) = \frac{1}{\sqrt{2}} \int_x^y e^{-\frac{t^2}{2}}\, dt \quad \forall x \le y$$

Der findes mange andre fordelinger med denne egenskab (de såkaldte uendelige delelige fordelinger, f.eks. eksponential fordelingen, Poisson fordelingen, gamma fordelingen, o.m.a.). Hver af disse fordelinger har deres egen *centrale grænseværdi lov*. F.eks. hvis $S_n = X_{1n} + \cdots + X_{nn}$ er en sum af n »små« tilfældige positive tal $X_{1n}, X_{2n}, \ldots, X_{nn}$, så gælder, under en række specifikke forudsætninger, at vi har:

$$\lim_{n\to\infty} P(x \le S_n \le y) = e^{-\lambda} \sum_{n=x}^{y} \frac{\lambda^n}{n!} \quad \forall 0 \le x \le y$$

og under en anden række specifikke forudsætninger, har vi

$$\lim_{n\to\infty} P(x \le S_n \le y) = \lambda \int_x^y e^{-\lambda t}\, dt \quad \forall 0 \le x \le y$$

Så normal fordelingen er ikke så »normal« endda – der er mange andre fordelinger, der fortjener navnet »normal fordelingen«. I sandsynlighedsregningen kan vi også bevise den såkaldte *iterede logaritme lov*, som løst sagt siger at i lange perioder (for meget lange observationsrækker) vil dine møntkast være meget langt væk fra en ligelig fordeling. Mere præcist, hvis S_n betegner antallet af »krone« i de første n kast, da gælder

$$\limsup_{n\to\infty} \frac{S_n - \frac{1}{2}n}{\sqrt{n \log\log n}} = 1 \quad \text{og} \quad \liminf_{n\to\infty} \frac{S_n - \frac{1}{2}n}{\sqrt{n \log\log n}} = -1$$

Faktisk er en »ligelig fordeling« undtagelsen snarere end reglen. Tilfældige hændelser opfører sig ganske anderledes end man forventer og ligger udenfor vores intuitions rækkevidde. I sandsynlighedsregningen kan vi beregne os til, hvorledes tilfældige hændelser opfører sig, men forstå vores resultater – det er en ganske anden sag. Efter et langt liv i sandsynlighedsregningen kan jeg stadigvæk blive overrasket, og jeg har lært at min intuition, selv efter jeg har ser de mange paradokser, som sandsynlighedsregningen er fuld af, er elendig, men jeg kan regne mig frem til korrekte resultater.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Formel for sansynlighed for at vinde i lotto

Gunnar Bruun

Kunne jeg få dig til at sende formel for sansynlighed for at vinde i lotto.

Svar

Sandsynligheden for at have 7 rigtige i lotto (ud af 36 mulige) er lig

$$\frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7}{36\cdot 35\cdot 34\cdot 33\cdot 32\cdot 31\cdot 30} = \frac{1}{8347680} \; .$$

I almindeligh er sandsynligheden for at have alle $(=x)$ rigtige i lotto (ud af n mulige) lig

$$\frac{1}{\binom{n}{x}} \quad \text{hvor} \quad \binom{n}{x} = \frac{n\cdot(n-1)1\cdot\ldots\cdot(n-x+1)}{1\cdot 2\cdot 3\cdot\ldots\cdot x}$$

De andre vinder kombinationer, f.eks. 6 rigtige og 1 forkert er mere bøvlede at skrive op, men du kan finde formlerne i Erling Andersens bog *Regn på din gevinst.*

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Et spil

Klaus Olsbjerg Jensen

Viden om i aftes fik mig til at tænke på et problem, jeg hørte for en del år siden, og som jeg aldrig har fundet en løsning på.

Det drejer sig om et simpelt spil. En spiller kan vælge mellem to kuverter. I hver kuvert er et pengebeløb, i den ene et dobbelt så stort beløb som i den anden. Spilleren vælger en kuvert, åbner den og finder 100 kr. Han får nu at vide at han gerne må vælge om hvis han vil.

Der er naturligvis ingen grund til at vælge om, sandsynligheden for at spilleren valgte kuverten med det store beløb er 50%.

Og dog, lad os regne lidt på det. I den anden kuvert er der enten 50 kr eller 200 kr. Så ved at skifte valg, vil spilleren i gennemsnit få $0{,}5\cdot 50\,\text{kr} + 0{,}5\cdot 200\,\text{kr} = 125\,\text{kr}$, altså 25% mere end hvis han ikke ændrede sit valg. Det må altså være en god strategi at vælge om.

Hvis beløbet havde været et andet, lad os sige X kr ville et omvalg i gennemsnit give spilleren $0{,}5\cdot 0{,}5\cdot X{,}\text{kr} + 0{,}5\cdot 2\cdot X\,\text{kr} = 1{,}25\cdot X\,\text{kr}$, altså stadig 25% mere.

Med andre ord, spilleren behøver ikke åbne kuverten. Ligegyldig hvilken kuvert han havde tænkt sig at vælge, ville det være smart at vælge den anden !!!

Hvor er fejlen ?

Svar

Problemet er, at sådan kan man ikke regne med (betingede) middelværdier. Lad os antage at beløbene er et helt antal kroner. Hvis beløbet i den kuvert du vælger er ulige; f.eks.: 201 kr, ved du at det er det lille beløb ($\frac{1}{2} \cdot 2001 = 100{,}5$, som ikke er et helt tal). Så ved at vælge om, er du sikker på at få det store beløb! Dette viser at dine udregninger ikke er korrekte.

Udregningerne ser mere eksotiske ud. Lad os nummerere konvolutterne 0 og 1. Lad X_0 være det tilfældige tal i konvolut nr. 0 og lad $X_1 = 2X_0$ være det tilfældige tal i konvolut nr. 1. Vi vælger nu en konvolut tilfældigt og uafhængigt af tallene X_0 og X_1; dvs. vi vølger et tal σ som enten er 0 eller 1, så at $P(\sigma = 0) = P(\sigma = 1) = \frac{1}{2}$ og σ er uafhængig af X_0 og X_1. Da er $\tau = 1 - \sigma$ nummeret på den ikke-valgte konvolut og X_τ er beløbet i den ikke-valgte konvolut. Vi åbner nu den valgte konvolut og ser at beløbet er lig x. Dette har sandsynlighed

$$P(X_\sigma = x) = \tfrac{1}{2} P(X_0 = x) + \tfrac{1}{2} P(X_0 = \tfrac{x}{2})$$

Lad A være den gennemsnitlige værdi af beløbet i den ikke-valgte konvolut. Da er A givet ved formlen

$$A = \frac{E(1_{\{X_\sigma = x\}} X_\tau)}{P(X_\sigma = x)}$$

Nu regner vi

$$\begin{aligned} E(1_{\{X_\sigma = x\}} X_\tau) &= E(1_{\{\sigma = 0, X_0 = x\}} X_1) + E(1_{\{\sigma = 1, X_1 = x\}} X_0) \\ &= 2x P(\sigma = 0, X_0 = x) + \tfrac{x}{2} P(\sigma = 1, X_1 = x) \\ &= x\big(P(X_0 = x) + \tfrac{1}{4} P(X_0 = \tfrac{x}{2})\big) \end{aligned}$$

Så A er givet ved formlen

$$E(X_\tau \mid X_\sigma = x) = x \frac{2P(X_0 = x) + \frac{1}{4} P(X_0 = \frac{x}{2})}{P(X_0 = x) + P(X_0 = \frac{x}{2})}$$

som i almindelighed hverken er større end x, mindre end x eller lig med x. For at afgøre om det er en fordel at vælge om, må du kende sandsynligheden for at det »lille« tal er lig x.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen

Spørgsmål til Jørgen Hoffmann-Jørgensen

Simon Olling Rebsdorf

Tak for en spændende udsendelse.

Jeg er selv cand.scient og nød hele baduljen. Men statistik var var aldrig min stærke side... Så kan man få lov at se formlen for stopstrategier i forb. med parkering, som så illustrativt blev præsenteret i programmet? (og en kort forklaring af de indgående parametre)

På forhånd tak!

Svar

Parkeringsproblemet: Som altid i sandsynlighedsteoretiske modeller er der em række underliggende antagelser. Her lyder de således: (a) Hændelserne »en P-plads er optaget« er uafhængige og har samme sandsynlighed, som er kendt, lad os sige p. (b) Antallet af P-pladser er kendt, lad os sige d. (c) Positionen af biografen er kendt, lad os sige ud for P-plads nummer n; det antages at $n \le d$, så der findes en P-plads udenfor biografen. (d) Der er opgivet en omkostning, lad os sige A, ved ikke at komme i biografen, så at A er større end afstanden fra sidste P-plads til biografen. (Dette krav betyder at du er villig til at tage sidste P-plads, selv om den ligger langt fra biografen.)

Den optimale P-strategi giver et bestemt tal, lad os sige r, som kan beregnes udfra størrelserne p, d, n og A, og den optimale strategi siger: »passér de første $r-1$ P-pladser, tag dernæst den første ledige P-plads derefter«. Den præcise formel for udregningen af den kritiske værdi r er eksotisk:

$$r = 1 \wedge \mathrm{ceil}\left(\frac{\log\{(A-d+n)p^d(1-p)+p^n(1-p)-p^d\}}{\log p}\right)$$

Her betyder $x \wedge y$ det største af tallene x og y, og ceil(x) er »gulvet i x'es lejlighed«; dvs. det største hele tal mindre end eller lig med x.

Hvis omkostningen A ved ikke at komme i biografen er meget stor, så er $r = 1$, og du skal vælge første ledige P-plads. Hvis $p \le \frac{1}{2}$, dvs. mindst havldelen af P-pladserne er ledige, da er $r = n$ og strategien sider: »Kør hen til biografen, og se om P-pladsen udenfor biografen er ledig. Hvis det er tilfældet tager du denne P-plads. Hvis ikke tager du den første ledige P-plads efter biografen, såfremt du kan finde en. I modsat fald opgiver du biografturen og går hjem«.

Med venlig hilsen
Jørgen Hoffmann-Jørgensen