Ragnar Frisch

SÆRTRYK AV
NORSK MATEMATISK
TIDSSKRIFT
1925

Fra Forfatteren

En ny metode i sandsynlighetsregningen.

Stanislas Millot: Théorie nouvelle de la probabilité des causes.

Gauthier-Villars, Paris 1925. 35 s. 5 fr.

Sandsynlighetsregningen er en av de faa omraader i matematikken som endnu ikke er blit «industrialiseret», hvor der endnu er rike frukter som kan høstes uten anvendelse av et høit specialiseret maskineri og en vidtdreven teknik. Vi ser et eksempel paa det i en nylig utkommet liten bok av den franske «capitaine de corvette en retraite» *Stanislas Millot.* Boken er et resumé av en række noter som forfatteren har faat fremlagt i Pariser-akademiet i aarene 1923—24. Den fremstiller et nyt princip for behandlingen av de opgaver som melder sig i forbindelse med sandsynlighetsregningens omvendings-problem: Paa grundlag av en empirisk bestemt relativ hyppighet at slutte bakover til den aprioriske sandsynlighet som har «ligget til grund» for forsøkene. Det emne som behandles er saaledes ikke som titelen synes at angi, teorien for aarsakers (hypotesers) sandsynlighet i sin fulde almindelighet, men den specielle teori som kommer til anvendelse, naar det erfaringsresultat man bygger paa, er en relativ hyppighet. Paa dette omraade har forfatteren med enkle analytiske hjælpemidler vundet resultater som efter anmelderens mening vil komme til at bli av grundlæggende betydning for den sandsynlighetsteoretiske vurdering av empirisk bestemte relative hyppigheter.

Som bekjendt gaar Bernoullis teorem[1]) ut paa, at hvis der med konstant sandsynlighet p gjøres s forsøk, saa eksisterer der

[1]) For at kunne bruke de betegnelser og uttryksmaater som er bedst kjendt for norske læsere, har jeg i det følgende ikke nøiagtig fulgt Millots betegnelser og heller ikke formuleret satsene paa nøiagtig samme maate som han.

en sandsynlighet eller sikkerhetsgrad $\Phi\ (x) \infty \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ for at talværdien av den iagttagne relative hyppighets avvikelse fra p, altsaa $|p' - p|$ skal være mindre end $x \sqrt{\frac{2pq}{s}}$, hvor $q = 1 - p$.

Omvendingen av Bernoullis teorem har i almindelighet været git følgende form:

Siden der eksisterer en sikkerhetsgrad $\Phi\ (x)$ for at $|p' - p| \lesseqgtr x \sqrt{\frac{2pq}{s}}$, saa maa det være tilnærmet rigtig at sige:

Der eksisterer en sikkerhetsgrad $\Phi(x)$ for at $|p' - p| \lesseqgtr x \sqrt{\frac{p'q'}{s}}$ hvor p' er den iagttagne relative hyppighet og $q' = 1 - p'$.

Det er let at finde eksempler hvor denne metode fører til uantagelige resultater. Hvis man betegner en sikkerhetsgrad som er lik eller større end $\Phi\ (3) = 0{,}999\,779\ldots$ som «næsten visshet», saa kan man, naar man gjør s forsøk med sandsynligheten p, være «næsten sikker paa» at $|p' - p| \leqq 3 \sqrt{\frac{2pq}{s}}$. La os anta at vi gjør $s = 400$ forsøk med sandsynlighet $p = 0{,}06$. Da vi har $3 \sqrt{\frac{2pq}{s}} =$ ca. $0{,}05$, kan vi være «næsten sikre paa» at faa en relativ hyppighet som ligger mellem $0{,}01$ og $0{,}11$. Vi maa altsaa være forberedt paa at faa en relativ hyppighet paa f. eks. $0{,}02$. Set at vi faar denne relative hyppighet og vil benytte omvendingsregelen saaledes som denne er formuleret ovenfor. Da vi har $3 \sqrt{\frac{2p'q'}{s}} =$ ca. $0{,}03$, saa skal vi ifølge regelen være «næsten sikre paa» at den sande værdi p av sandsynligheten ligger mellem $0{,}02 - 0{,}03 = \div\ 0{,}01$ og $0{,}02 + 0{,}03 = 0{,}05$. Den nedre av disse grænser har ingen mening, da den er negativ, og den øvre grænse er for lav. Den sande værdi av p var nemlig $0{,}06$ og falder altsaa utenfor intervallet. Slike urimelige resultater fører den opstillede omvendingsregel til, naar antallet av forsøk ikke er meget stort, altsaa netop i de tilfælde da sandsynlighetsregningens anvendelse er mest paakrævet.

For at undgaa disse urimeligheter utvikler Millot en metode ved hvilken det blir mulig at foreta omvendingen av Bernoullis teorem *uten at bruke den tilnærmelse som bestaar i at man* erstatter p av p' i uttrykket $\sqrt{\frac{2pq}{s}}$.

Hans ræsonnement er ganske enkelt følgende:

Der eksisterer en sikkerhetsgrad $\Phi\ (x)$ for at

$$|p - p'| \overline{\gtrless} x \sqrt{\frac{2pq}{s}}.$$

d. v. s. for at

$$p^2 \left(1 + \frac{2\,x^2}{s}\right) - 2\,p\left(p' + \frac{x^2}{s}\right) + p'^{\,2} \overline{\gtrless} 0$$

Denne betingelse vil være opfyldt, naar p har en værdi som ligger mellem ordinatværdien paa øvre og nedre gren av den kurve, hvis ligning i retvinklede koordinater (x, p) er

$$(1) \qquad p = \frac{sp' + x^2}{s + 2\,x^2} \pm \frac{x}{s + 2\,x^2}\sqrt{x^2 + 2\,sp'\,(1 - p')}.$$

Problemet er derfor reduseret til det at studere forløpet av disse to kurvegrener. Man overbeviser sig let om at kurvegrenene møtes i et dobbeltpunkt, der ligger paa p-aksen med ordinat $p = p'$. For positive x ligger kurvegrenene i sin helhet mellem de rette linjer $p = 1$ og $p = 0$. Disse linjer er aymptoter for henholdsvis øvre og nedre gren. Arealet mellem de to grener kalder Millot den til den relative hyppighet $p' = \frac{m}{s}$ hørende sandsynlighets-sone.

Ved en enkel betragtning over sandsynlighets-sonen kan man opstille satsen:

Hvis der foreligger en empirisk bestemt relativ hyppighet $p' = \frac{m}{s}$, saa eksisterer der en sikkerhetsgrad $\Phi\ (x)$ for at den sande værdi p av sandsynligheten ligger mellem de to grænser som er git ved (1).

Ved at benytte denne sats finder man for det ovennævnte taleksempel at man kan være «næsten sikker paa» at den sande værdi p ligger mellem 0,005 og 0,076. Indenfor disse grænser ligger ogsaa ganske rigtig den værdi 0,06 som man i eksemplet gik ut fra.

Det er interessant at lægge mærke til at den gamle metode kan vises at være ensbetydende med at man erstatter de to kurvegrener av deres tangenter i dobbeltpunktet $(0, p')$. Nedre grænse for den sandsynlighets-sone som derved fremkommer, blir for tilstrækkelig store x negativ, et tilfælde som netop er illustrert i det behandlede taleksempel.

Ved en videre analyse av sandsynlighets-sonen løser saa Millot de øvrige klassiske omvendingsopgaver. F. eks. følgende opgave:

Der foreligger en empirisk bestemt relativ hyppighet $p' = \frac{m}{s}$. La p være den sande værdi av sandsynligheten, og la p_0 være et visst fast opgit tal mellem 0 og 1. Hvad er sikkerhetsgraden for at p skal ligge indenfor et spillerum (omkring p'), der er saa stort at det netop saavidt inkluderer p_0? Sættes

$$x_0 = \left| \frac{p_0 - p'}{\sqrt{\frac{2\, p_0\, (1 - p_0)}{s}}} \right| \tag{2}$$

saa er den søkte sikkerhetsgrad $\Phi\ (x_0)$.

Videre: Sikkerhetsgraden for at p ligger paa samme side av p_0 som den empiriske værdi p' er $^1/_2 \left\{1 + \Phi(x_0)\right\}$ (3)

Som en anvendelse av de utviklede sætninger behandles bl. a. den bekjendte opgave som Laplace stillet sig: Av samtlige 493 472 fødsler registrert i Paris i aarene 1745—70 var 251 527 guttefødsler. Bestem den sikkerhetsgrad som tilkommer følgende dom: «Sandsynligheten for at en nyfødt pariser skal være gut er større end $^1/_2$». Man finder en sikkerhetsgrad hvor de første 40 decimaler er 9'ere.

Til bruk under teoriens anvendelse gis et nomogram ved hjælp av hvilket x_0 kan beregnes.

De problemer som reiser sig i forbindelse med analysen av *to* empirisk bestemte relative hyppigheter $p' = \frac{m'}{s'}$ og $p'' = \frac{m''}{s''}$ behandler han ved at indføre størrelsen

$$\bar{p} = \frac{s'\, p' - s''\, p''}{s' - s''} \pm \frac{\sqrt{s'\, s''}}{s' - s''}\, (p' - p'') \tag{4 a}$$

Hvor tegnet vælges slik at man benytter uttrykkets største værdi, naar $(p' - p'')$ og $(s'\, p' - s''\, p'')$ har motsat tegn, men mindste værdi, naar disse to størrelser har samme tegn.

Videre indføres størrelsen

(4 b) $$\bar{x} = \frac{\bar{p} - p'}{\sqrt{\frac{2\,\bar{p}\,(1-\bar{p})}{s'}}} = \frac{\bar{p} - p''}{\sqrt{\frac{2\,\bar{p}\,(1-\bar{p})}{s''}}}$$

Ved hjælp av disse størrelser formuleres forskjellige sætninger, f. eks. denne:

Hvis der foreligger de empirisk bestemte relative hyppigheter $p' = \frac{m'}{s'}$ og $p'' = \frac{m''}{s''}$, og de tilsvarende sande værdier av sandsynlighetene betegnes p_1 og p_2, saa eksisterer der en sikkerhetsgrad

(5) $$P = {}^1/_2 \left(1 + \Phi\,(\bar{x}) \right)$$

for at størrelsesrækkefølgen mellem p_1 og p_2 er den samme som størrelsesrækkefølgen mellem p' og p''.

I sidste kapitel gaar han ind paa det tilfælde, da de enkelte forsøk indenfor serien ikke længer er indbyrdes uavhængige, da altsaa forutsætningene for anvendelsen av Bernoullis teorem i dets oprindelige form ikke længer er tilstede. Istedenfor uten videre at gaa ut fra at middelfeilen paa p' er lik $\sqrt{\frac{p\,(1-p)}{s}}$ gaar han ut fra at middelfeilen er lik $k\sqrt{\frac{p\,(1-p)}{s}}$, hvor k er en faktor (eventuelt funktion av s) som blir nærmere at bestemme efter arten av vedkommende forsøk. Herimot kan der selvfølgelig ingen indvending gjøres. Under de videre betragtninger over faktoren k benyttes imidlertid en uttryksmaate som maa siges at være litt uheldig, idet han betegner k som «det omvendte av det man i feilteorien kalder presisjonen». Dette kan ikke egentlig siges at stemme med almindelig sprogbruk. Det omvendte av presisjonen er jo hele uttrykket $k\sqrt{\frac{2\,p\,(1-p)}{s}}$, mens faktoren k paa en viss maate blir et uttryk for korrelationen mellem de enkelte forsøk i serien. Der kan ogsaa gjøres den indvending at faktoren k blir behandlet som om den skulde være en størrelse, der i hvert enkelt tilfælde maa gjøres til gjenstand for et *skjønsmæssig* anslag. Paa dette punkt vilde det været ønskelig om teorien i større utstrækning hadde utnyttet resultatene fra de nyere undersøkelser paa omraadet. Disse gaar netop ut paa at finde uttryk for middelfeilen paa et gjennemsnit av iagttagelser som ikke er indbyrdes uavhængige. Middelfeilen paa et gjentagelsestal ved ikke uavhængige forsøk falder ind som specialtilfælde herunder. Disse spørsmaal er blit behandlet bl. a. av professor Tschuprow.

Man bør ogsaa være opmerksom paa at under teorien for uavhængige forsøk er formel (3) utledet under følgende forutsætning: Det er *aposteriori* like stor sandsynlighet for relationen $p > p'$ som for relationen $p < p'$. Denne forutsætning holder stik som en første tilnærmelse, men den er ikke eksakt. Allerede den omstændighet[1]) at det *apriori* er større sandsynlighet for relationen $p > p'$ end for relationen $p < p'$, naar $p < {}^1/_2$ og omvendt naar $p > {}^1/_2$ (forutsat at sp er et helt tal), peker i retning av at Millots forutsætninger paa dette punkt ikke holder stik andet end som en tilnærmelse. Dette bekræftes ogsaa ved en nærmere analyse. Forutsætter man nemlig at sp er et helt tal, altsaa at p har en av de diskrete værdier $\frac{k}{s}$ $(k = 0, 1 \ldots s)$ saa viser det sig at spørsmaalet om relationene $p \gtrless p'$ er like sandsynlige, reduseres til spørsmaalet om uttrykkene

$$\sum_{k=0}^{m-1} k^m \, (s - k)^{s-m} \quad \text{og} \quad \sum_{k=m+1}^{s} k^m \, (s - k)^{s-m}$$

er like. Man overbeviser sig let ved taleksempler om at dette ikke er tilfælde i sin almindelighet. Gjør man ingen forutsætninger om at sp er et helt tal, faar man at foreta en sammenligning mellem integralene

$$\int_0^{\frac{m}{s}} x^m \, (1-x)^{s-m} \, dx \quad \text{og} \quad \int_{\frac{m}{s}}^{1} x^m \, (1 - x)^{s-m} \, dx$$

Disse integraler kan uttrykkes ved visse specielle værdier av den ukomplette B-funktion som kan sammenlignes ved en anvendelse av Simmons teorem. Ved at gjennemføre analysen vil man formodentlig kunne paavise ikke bare at Millots forutsætninger paa dette punkt kun er tilnærmet rigtige, hvad allerede enkle taleksempler viser, men man vil formodentlig ogsaa kunne opstille positive regler for hvilke av de to relationer $p \gtrless p'$ der er den

---

[1]) Simmons teorem. Skandinavisk Aktuarietidskrift. 1924, p. 153. «Solution d'un problème du calcul des probabilités».

sandsynligste. Sandsynligvis vil man finde at under visse almindelige forutsætninger er det aposteriori større sandsynlighet for relationen $p > p'$ end for relationen $p < p'$, naar $p' < {}^1/_2$. Omvendt naar $p' > {}^1/_2$.

Den unøiagtighet som paa dette punkt klæber ved Millots metode er imidlertid av underordnet betydning sammenlignet med de utvilsomme fordeler metoden medfører. Naar det gjælder at vurdere sikkerhetsgraden av en relativ hyppighet som er fremgaat av indbyrdes uavhængige forsøk, betegner metoden et væsentlig fremskridt. For fremtiden vil ingen lærebok i sandsynlighetsregning være fuldstændig, hvis den ikke indeholder et avsnit om den nye metode.

Ragnar Frisch.