

NOTAT

Til Christian Westrup Jensen

Vedr. Multiple choice scoringsalgoritme

Fra Bjarne Kjær Ersbøll

5. december 2013
be

Forsøg på en enkel forklaring på den implementerede scoringsalgoritme

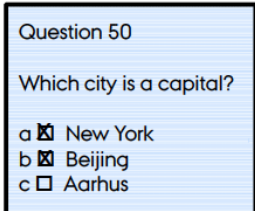
Scoringsalgoritmen fremgår af Michael I. Schwartzbach's slides:

The score for a single question is $S(k,a,c)$ where:

k is the number of choices

a is the number of **X** marks written by the student

$c \in \{0,1\}$ denotes whether the correct answer is given

score() = $S(3,2,1) = ???$

$$S_{\text{axiom}}(k,a,c) = \begin{cases} 0 & \text{if } a = 0 \text{ or } k = a \\ \log(k/a) & \text{if } k > a > 0 \text{ and } c = 1 \\ -a/(k-a) \log(k/a) & \text{if } k > a > 0 \text{ and } c = 0 \end{cases}$$

Tilfældet Indledende Statistik

På DTU Compute har vi i omtrent 25 år haft multiple choice eksamen i det indledende statistikkursus. Formen har altid været med 30 enkeltspørgsmål, hver med 5 svarmuligheder med præcis 1 korrekt svar og 4 forkerte svar. Der må sættes et og kun et kryds. (Note: dette passer ind som en restriktiv variant af Michael I. Schwartzbach's mere generelle scoringsalgoritme.)

I det tilfælde er det relativt enkelt at indse, at +4 point for et korrekt svar og -1 point for et forkert svar og 0 point for "ved ikke" (=ingen svar) giver i gennemsnit 0 point, hvis den studerende forsøger at "tipse".

For den slags multiple choice eksamener er for den implementerede formel:

- antal svarmuligheder: $k=5$
- antal krydser sat af den studerende: $a=1$ (men $a=0$ for "ved ikke")
- ved korrekt svar er: $c=1$; ved forkert svar er: $c=0$

Scoren for et korrekt svar vil da være: $\log(k/a)=\log(5/1)=\log(5)= 1.6094$

Scoren for et forkert svar er: $-a/(k-a)\log(k/a)=-1/(5-1)\log(5/1)=-1/4\log(5)= -0.4024$

Scoren for "ved ikke" er: 0

Det noteres at der kun er en proportionalitetsfaktor til forskel fra de +4 og -1 point vist ovenfor. Metoden kan derfor uden videre benyttes til vores kursus i indledende statistik.

(For god ordens skyld skal nævnes, at der ofte har været brugt: +5 point for rigtigt svar og -1 point for forkert – dette giver blot en forventningsværdi på +1 i stedet for 0 for et spørgsmål, eller +30 point for hele sættet i stedet for 0.)

Nogle andre egenskaber ved scoringsalgoritmen

Mere end et kryds

Imidlertid tillader metoden mere fleksibilitet end vi bruger lige nu i indledende statistik fx: den studerende må sætte mere end et kryds. Intuitivt skal det selvfølgelig ikke give helt så høj en score, hvis et af krydserne er sat det rigtige sted, mens 2 forkerte satte krydser nok bør straffes lidt hårdere end ved kun et forkert sat kryds (a la halvgarderinger i tips, som jo også er dyrere):

- Scoren for et korrekt svar (et af krydserne er sat rigtigt) vil da være: $\log(k/a)=\log(5/2)= 0.9163$ altså positivt, men ikke så godt, som hvis man svarer rigtigt ved kun at sætte et enkelt kryds.
- Scoren for et forkert svar (begge krydser er sat forkert) er: $-a/(k-a)\log(k/a)=-2/(5-2)\log(5/2)=-2/3\log(5/2)= -0.6109$ altså en lidt større straf for at have demonstreret, at man ikke kender det rigtige svar nemlig med hele 2 forkerte krydser.
- Scoren for "ved ikke" (ingen krydser eller krydser i alle felter) er selvfølgelig stadig: 0

Forskelligt antal svarmuligheder ved forskellige spørgsmål

Der er også mulighed for at blande spørgsmål med fx 5 svarmuligheder og andre spørgsmål med 4, 3 eller 2 svarmuligheder. Nogle gange er de relevante svarmuligheder på et spørgsmål fx "ja" eller "nej". I det tilfælde vil 2 svarmuligheder være relevante. Det snedige ved metoden er, at den tager hensyn til, at det er lettere at "tipse" sig frem til det rigtige svar ved et ja/nej spørgsmål:

- Scoren for et korrekt svar vil da være: $\log(k/a)=\log(2/1)=\log(2)= 0.6931$ (tæller altså positivt, men ikke så meget, som ved 5 svarmuligheder)
- Scoren for et forkert svar er: $-a/(k-a)\log(k/a)=-1/(2-1)\log(2/1)=-\log(2)= -0.6931$ (tæller altså negativt, og mere end ved 5 svarmuligheder)
- Scoren for "ved ikke" (0 krydser eller to krydser) er: 0

At det scoringsmæssigt giver mening at kunne have spørgsmål med forskelligt antal svar kan illustreres ved at have et spørgsmål bestående af to underspørgsmål, som hver for sig kunne besvares med ja/nej. Det kan formuleres som to spørgsmål med svarmulighederne ja/nej eller et spørgsmål med fire svarmuligheder: ja+ja, ja+nej, nej+ja, nej+nej.

- Scoren for et korrekt svar vil da være: $\log(k/a)=\log(4/1)=\log(4)= 1.3863$ (altså præcis det dobbelte af et enkelt korrekt ja/nej spørgsmål)
- Scoren for et forkert svar er: $-a/(k-a)\log(k/a)=-1/(4-1)\log(4/1)=-1/3\log(4)= -0.4621$ (rigtigt underspørgsmål 1 (+0.6931) og forkert underspørgsmål 2 (-0.6931) giver 0, forkert underspørgsmål 1 (-0.6931) og rigtigt underspørgsmål 2 (+0.6931) giver selvfølgelig også 0, mens forkert underspørgsmål 1 (-0.6931) og forkert underspørgsmål 2 (-0.6931) giver -1.3863. Da der var 3 forkerte svarmuligheder giver det i snit: $(0+0-1.3863)/3=-0.4621$)
- Scoren for "ved ikke" (0 krydser eller 4 krydser) er: 0

Det er altså ligegyldigt (eksakt for rigtige svar, og i snit for forkerte svar), om man laver to spørgsmål med ja/nej svarmuligheder, eller et spørgsmål med ja+ja, ja+nej, nej+ja, nej+nej svarmuligheder.