Produkt og marked - betinget sandsynlighed

Rasmus Waagepetersen
Institut for Matematiske Fag Aalborg Universitet

February 21, 2022

Tænkeboks opgave i Ingeniøren

Se webside https://ing.dk/artikel/taenkeboks-sandsynligheden-fejl-ved-positiv-test-221355

"I denne uge kommer opgaven igen fra SDU, Mads Clausen Instituttet i Sønderborg, og her skal kvalitetsingeniøren i arbejde:"

Opgave 5: I en produktionsvirksomhed udtages med jævne mellemrum et eksemplar af et produkt, der sendes til en automatiseret test for fejl.

Erfaringsmæssigt vides, at 1 ud af 1000 produkter er defekt. Den automatiserede test er (naturligvis) ikke helt fejlfri selv, idet den vil vise en falsk positiv i 2% af tilfældene og en falsk negativ i 4% af tilfældene.

Hvad er sandsynligheden for, at et produkt er defekt, når den automatiserede test er positiv?

Svar under kommentarer (i alt 26):

1) 4,58%. Omend man kan undre sig over formålet med en så ringe test

2) Vi benytter loven om betingede sandsynligheder på situationen hvor et tilfældigt produkt testes. Lad F, D være begivenhederne at produktet er henholdsvis fungerende og defekt. Lad yderligere TP og TN være begivenhederne, at testen viser positiv eller negativ, henholdsvis. Da er P(D — TP) = P(D og TP)/P(TP)=P(D og TP)/(P(TP og D) + P(TP og F)) = 1/(1+P(TP og F)/P(TP og D)). Vi har P(TP og F) = (999/1000)(96/100) P(TP og D) = (1/1000)(2/100) således at P(TP og F)/P(TP og D) = 999*48=49592. Dermed er svaret 1/49593.

3) Kim Bygum har ret.
Her en hurtig pædagogisk forklaring.
Hvis vi sender 100000 igennem testprocessen vil 100 være defekte. Her vil 96 af de defekte give et positivt testresultat og 4 vil give falsk negativ. Af de resterende 99900 som ikke er defekte vil 2 procent altså 1998 give falsk positiv og resten 97902 give negativ.
Det samlede antal positive bliver 96+1998=2094
Og procentdelen af de positive testresultater, der rent faktisk er defekte bliver 96/2094 * 100% = 4,58 %

4) Nej, hør nu! 2% af 1000 er 20, som er antallet af falske positiver. En (anden) ud af de tusind er en 'rigtig' positiv. I alt 21 'positiv-visninger'. Så sandsynligheden for at en enhed er defekt når der vises 'positiv', er 1 ud af 21, altså 4,76%

Betinget sandsynlighed

Lad A og B angive to hændelser.

Den betingede sandsynlighed for A givet B noteres $P(A|B)$ og er sandsynligheden for at A indtræffer når vi ved, at B er indtruffet.

Eksempel (kast med terning): antag, at vi har fået informationen B, at et terninge kast viser mere end eller lig 3. I lyset af dette, hvad er så sandsynligheden for, at kastet er 4 (hændelsen A)?

De mulige udfald er nu $B = \{3, 4, 5, 6\}$ og alle 4 udfald er stadig lige sandsynlige. Dvs. sandsynligheden for A givet B er $1/4$.

Definition: den betingede sandsynlighed for en hændelse A givet en hændelse B defineres som

$$P(A|B) = \frac{P(A \text{ og } B)}{P(B)}$$

Terningekastet: $A = \{4\}$ og $B = \{3, 4, 5, 6\}$.
$P(A \text{ og } B) = P(\{4\}) = 1/6$. $P(B) = 4/6$. $P(A|B) = 1/4$.

Kvalitetskontrol

I opgaven får vi at vide, at $P(D) = 1/1000$, hvor D er hændelsen, at et objekt er defekt.

Yderligere får vi at vide, at sandsynligheden for falsk positiv er 2%. Dette er den betingede sandsynlighed for at testen er positiv givet at produktet er OK (ikke D).

NB: positiv test betyder at testen siger, at produktet er defekt ! (jf. "negativt" at være covid-19 eller HIV-positiv)

Endeligt er sandsynligheden for falsk negativ 4%. Det betyder, at der er 4% sandsynlighed for at testen siger, at produktet er OK givet at produktet er defekt (D).

Hvis vi lader D_{test} være hændelsen, at testen er positiv, kan de to ovenstående oplysninger formuleres som $P(D_{\text{test}}|\text{ikke } D) = 2\%$ og $P(\text{ikke } D_{\text{test}}|D) = 4\%$.

Sandsynligheden, som opgaven efterspørger er $P(D|D_{\text{test}})$!

Altså, givet at et produkt er testet positiv (defekt), hvad er så sandsynligheden for, at det reelt er defekt ?

Vi har brug for at regne "modsat" med betingede sandsynligheder.

Pædagogisk forklaring :)

Kvalitetskontrollen kan illustreres vha. en tabel:

	D_{test} (positiv)	ikke D_{test} (negativ)
D	D og D_{test}	D og ikke D_{test} (falsk negativ)
ikke D	ikke D og D_{test} (falsk positiv)	ikke D og ikke D_{test}

Forventede antal med 100.000 produkter:

	D_{test} (positiv)	ikke D_{test} (negativ)	
D	96	4	100
ikke D	1998	97902	99900
	2094	97906	100000

Vi ser nu:

$$P(D|D_{\text{test}}) = \frac{96}{2096} = 0.0458$$

I stedet for antal kan vi se på sandsynligheder:

	D_{test} (Positiv)	ikke D_{test} (Negativ)	
D	0.00096	0.00004	0.001
ikke D	0.01998	0.97902	0.999
	0.02094	0.97906	1

Bemærk

$$P(D \text{ og } D_{\text{test}}) = P(D_{\text{test}}|D)P(D) = 0.96 \cdot 0.001 = 0.00096$$

og

$$P(D|D_{\text{test}}) = \frac{P(D \text{ og } D_{\text{test}})}{P(D_{\text{test}})} = \frac{P(D_{\text{test}}|D)P(D)}{P(D_{\text{test}})} = \frac{0.00096}{0.02094} = 0.0458$$

Dette er eksempel på Bayes sætning !

Bayes sætning

Antag, at vi kender $P(A|B)$, $P(A)$ og $P(B)$. Så er

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Hvis vi ikke kender A men derimod både $P(A|B)$ og $P(A|\text{ikke } B)$ kan vi udregne $P(A)$ som

$$\begin{aligned} P(A) &= P(A \text{ og } B) + P(A \text{ og ikke } B) \\ &= P(A|B)P(B) + P(A|\text{ikke } B)P(\text{ikke } B) \end{aligned}$$

Eksempel (terningekast): A kast giver 4 og B kast større eller lig 3. $P(A) = 1/6$ og $P(B) = 4/6$. $P(A|B) = 1/4$.

$$P(B|A) = \frac{(1/4)(4/6)}{1/6} = 1$$

(selvfølgelig)

Beregning af $P(D|D_{\text{test}})$ vha. Bayes:

$$P(D|D_{\text{test}}) = \frac{P(D_{\text{test}}|D)P(D)}{P(D_{\text{test}})}$$

Tæller: $P(D_{\text{test}}|D) = 1 - P(\text{ikke } D_{\text{test}}|D) = 0.96$ dvs.
$P(D_{\text{test}}|D)P(D) = 0.96 \cdot 0.001 = 0.00096$.

Nævner:
$P(D_{\text{test}}) = P(D_{\text{test}}|D)P(D) + P(D_{\text{test}}|\text{ikke } D)P(\text{ikke } D) =$
$0.96 \cdot 0.001 + 0.02 \cdot 0.999 = 0.02094$

Dvs. $P(D|D_{\text{test}}) = 0.00096/0.02094 = 0.0458$!

(NB: samme udregning som "pædagogisk" forklaring. Blot med sandsynligheder i stedet for antal)

Hvad siger dette om konsekvenserne af kvalitetskontrollen ?

Opgaver

1. Lad A være hændelsen, at et terningekast er ≥ 4 og B hændelsen, at kastet er ≤ 5. Hvad er $P(A|B)$ og $P(B|A)$?
2. (udvidelse af opgaven fra Mads Clausen instituttet) Beregn $P(\text{ikke } D|\text{ikke } D_{\text{test}})$ - altså sandsynligheden for, at et produkt ikke er defekt, givet at testen er negativ. Hvilken konsekvens har denne sandsynlighed for aftagerne af produktet (som kun får produkter med negativ test).

3. Politiet har Hans og Anders i kikkerten som mistænkte for et bankrøveri, hvor der kun var en gerningsmand. Som udgangspunkt er Hans og Anders lige sandsynlige som gerningsmænd, dvs. $P(H) = P(A) = 1/2$, hvor H er hændelsen, at Hans er gerningsmanden og A er hændelsen, at Anders er gerningsmanden.

   Der er indsamlet en DNA-profil D, men pga. usikkerheder i analysen af DNA-materialet, er der ikke en entydig sammenhæng mellem gerningsmand og DNA-profil

   Lad $P(D|H)$ være sandsynligheden for den indsamlede DNA-profil givet at Hans er gerningsmanden og tilsvarende $P(D|A)$ hvis Anders er gerningsmanden. Retsgenetikere beregner, at $P(D|H) = 0.2$ og $P(D|A)$ er 0.6.

   Hvad er nu $P(H|D)$, altså sandsynligheden for at Hans er gerningsmanden givet bevismaterialet i form af DNA-profilen ?