Bayes theorem

23.10.2005

GE02 day 2 part 2

Yurii Auchenko

Erasmus MC Rotterdam

Problem

P(M)=0.05, P(D|MM)=1.0, P(D|MN)=0.7 P(D|NN)=0.03 (same as problem before)

- if we observe an ill person, what is the probability it would have genotype MM, MN or NN?
- ...to put it formally, what are the genotypic probabilities given a person is ill, P(MM|D), P(MN|D) and P(NN|D)?
- These are the probabilites of the genotypes in a “population” of ill people!

Solution

Probability of disease, P(D) = 0.0961

This probability was made of three components:

- 0.0025 (these ill from MM) + 0.0665 (from MN) + 0.0271 (from NN) = 0.0961

Thus, the proportion of

- MM is 0.0025/0.0961 = 0.026 = 2.6%
- MN is 0.0665/0.0961 = 0.6922 = 69.22%
- NN is 0.0271/0.0961 = 0.2818 = 28.18%

Bayes' formula

We were following the schema

P(M)	0,05		P(D&g)=	P(g|D)=
g	P(g)	P(D|g)	P(g)*P(D|g)	P(g)*P(D|g)/P(D)
MM	0,0025	1,0000	0,0025	0,0260
MN	0,0950	0,7000	0,0665	0,6922
NN	0,9025	0,0300	0,0271	0,2818
		P(D)=	0,0961	

And the computations were done using the formula

$$P(g \mid D) = \frac{P(D \mid g)P(g)}{P(D)} = \frac{P(D \mid g)P(g)}{\sum_{g=MM,MD,DD} P(D \mid g)P(g)}$$

Total probability and Bayes' formulas

Two sets of events are considered:

- "Hypothesis" H_i for which *a prioi* probabilities, $P(H_i)$ are known. E.g. genotypes were "hypotheses" in our example. These hypotheses must be mutually exclusive.
- Event(s) of interest, A, e.g. disease. For this event, conditional probabilites given hypotheses, $P(A|H_i)$

Total probability & Bayes formulae

Total probability (of event A)

$$P(A) = \sum_i P(A|H_i)P(H_i)$$

Probability of hypothesis H_i, given A

$$P(H_i | A) = \frac{P(A|H_i)P(H_i)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum_i P(A|H_i)P(H_i)}$$

Problem

- You pick up a bowl at random, and then pick up a cookie at random. The cookie turns out to be a plain one.
- Use Bayes's formula to find out what is the probability that you picked the cookie out of bowl #1

Solution

- H_1 – bowl number 1
- H_2 – bowl number 2
- A – plain cookie
- $P(H_1) = P(H_2) = ½$
- $P(A| H_1) = ¾$
- $P(A| H_2) = ½$

$$P(H_1 | A) = \frac{P(A|H_1)P(H_1)}{P(A)} = \frac{P(A|H_1)P(H_1)}{\sum_{i=1,2} P(A|H_i)P(H_i)}$$

= (¾ ½) / (¾ ½ + ½ ½) = (3/8) / (5/8) = 3/5

Problem

- In a population, the frequency of obese people is 25%, overweight is observed in 40% and normalweight people have frequency of 35%. The frequency of hypertension in these groups is 45, 30 and 20%, respectively
  - What is the total frequency of hypertension in the population?
  - If a random person is hypertensive, what is the best guess about his (her) weight?
  - If a random person is not hypertensive, what is the best guess about his (her) weight?

Solution

- Denote
  - H1=obese, H2=overweight and H3=normal
  - A = hypertensive, B=not hypertensive
- Probabilities
  - $P(H1)=0.25$, $P(H2)=0.4$ and $P(H3)=0.35$
  - $P(A|H1)=0.45$, $P(A|H2)=0.3$ and $P(A|H3)=0.2$
  - $P(B|H1)=1 - P(A|H1) = 0.55$, $P(B|H2)=0.7$ and $P(B|H3)=0.8$

Solution: frequency of hypertension

- Probabilities
  - P(H1)=0.25, P(H2)=0.4 and P(H3)=0.35
  - P(A|H1)=0.45, P(A|H2)=0.3 and P(A|H3)=0.2

$$P(A) = \sum_{i=1,2,3} P(A|H_i)\,P(H_i)$$

$$P(A|H_1)\,P(H_1) + P(A|H_2)\,P(H_2) + P(A|H_3)\,P(H_3)$$

$$0.25\ 0.45 + 0.4\ 0.3 + 0.35\ 0.2 = 0.3$$

Solution: weight groups frequencies

- Probabilities
  - P(H1)=0.25, P(H2)=0.4 and P(H3)=0.35
  - P(A|H1)=0.45, P(A|H2)=0.3 and P(A|H3)=0.2

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

$$P(H_1|A)=\frac{P(A|H_1)P(H_1)}{P(A)}=\frac{0.25\,0.45}{0.3}=0.37$$

$$P(H_2|A)=\frac{P(A|H_2)P(H_2)}{P(A)}=\frac{0.4\,0.3}{0.3}=0.4$$

$$P(H_3|A)=\frac{P(A|H_3)P(H_3)}{P(A)}=\frac{0.35\,0.2}{0.3}=0.23$$

