AMS131 23 Apr 19

(see syllabus)

(see syllabus)

credit card screening case study (continued) ①

see document camera notes 18 Apr 19

we know

$P(B) = 0.01$

$P(\ominus \mid G) = 0.97$

$P(\oplus \mid B) = 0.98$

we want $P(B \mid \oplus)$

truth (unknown) | system says (data)

only 2 possible truth states (B, G)

solution method 1: 2×2 table

method 2 | Bayes's Theorem in odds form

earlier we showed that

(likelihood ratio)

$$\frac{P(B \mid \oplus)}{P(\text{not } B \mid \oplus)} = \left[\frac{P(B)}{P(\text{not } B)}\right] \cdot \left[\frac{P(\oplus \mid B)}{P(\oplus \mid \text{not } B)}\right] \quad (*)$$

G

G

$$\begin{pmatrix}\text{posterior} \\ \text{odds in} \\ \text{favor of } B\end{pmatrix} = \begin{pmatrix}\text{prior} \\ \text{odds} \\ \text{in favor} \\ \text{of } B\end{pmatrix} \cdot \begin{pmatrix}\text{Bayes factor} \\ \text{in favor of } B\end{pmatrix}$$

excellent fact: we know all the probabilities on the right-hand side of $\circledast$

$P(B) = 0.01$

$P(\text{not } B) = P(G) = 0.99$

$P(\oplus \mid B) = 0.98$

$P(\oplus \mid \text{not } B) =$
$1 - P(\ominus \mid \text{not } B)$
$= 1 - P(\ominus \mid G)$
$= 1 - 0.97$
$= 0.03$

so let's compute:

$$\frac{P(B \mid \oplus)}{P(\text{not } B \mid \oplus)} = \left(\frac{0.01}{0.99}\right)\left(\frac{0.98}{0.03}\right)$$

(99 to 1 prior odds against B) (98 to 3 odds in favor of B)

$$= \frac{98}{(99)(3)} = \frac{98}{297}$$

(posterior odds in favor of B)

recall $o = \frac{p}{1-p}$

odds ... probability

from which

$$p = \frac{o}{1+o} = \frac{\frac{98}{297}}{1 + \frac{98}{297}} = \frac{98}{98+297} = \frac{98}{395} \doteq 0.25$$

as before

method 3

Bayes's Theorem in probability form ③

$$P(B \mid \oplus) = \frac{\overset{0.01}{P(B)}\ \overset{0.98}{P(\oplus \mid B)}}{P(\oplus)}$$

annoying denominator (normalizing constant)

? prediction

how compute $P(\oplus)$?

(truth) {B, G} (unknown)

data (observable)

(data) {⊕, ⊖}

$P(B \mid \oplus)$

truth (unknown) data

search for truth on basis of data

(statistical) inference

(statistical) inference about data

↔ prediction

$P(\oplus) = ?$

I don't know much at all about ⊕ by itself; if I knew B or G, bingo

unknown (true) state of world may well be unobservable

Y (want to predict) (hard)

new creative laziness idea: (D. V. Lindley)

when Y is hard, [get help] [extend the conversation]

find some other aspect of world X upon which Y depends, & predict Y in 2 stages: $\left\{ \begin{matrix} X \\ (Y \mid X) \end{matrix} \right\}$ indirect

direct

data (Y) other info (X): $\left\{ \begin{matrix} G \\ B \end{matrix} \right\}$

$$P(\oplus) = P(\oplus \text{ and } G) + P(\oplus \text{ and } B)$$

or

$$= P[(\oplus \text{ and } G) \text{ or } (\oplus \text{ and } B)]$$

were partitioning over the truth

$$= P(G)\,P(\oplus \mid G) + P(B)\,P(\oplus \mid B)$$

$$\overset{\text{I}}{P(G)}\cdot\overset{(\text{II}\mid\text{I})}{P(\oplus\mid G)} + \overset{\text{I}}{P(B)}\cdot\overset{(\text{II}\mid\text{I})}{P(\oplus\mid B)}$$

$$P(\oplus) = P(G)\cdot P(\oplus\mid G) + P(B)\cdot P(\oplus\mid B)$$

$$= (0.99)\,[1 - \underset{0.97}{P(\ominus\mid G)}]\,(0.01)\,(0.98)$$

so finally

$$= (0.99)(0.03) + (0.01)(0.98)$$

$$(0.0297) + (0.0098) = 0.0395$$

$$P(B\mid\oplus) = \frac{P(B)\,P(\oplus\mid B)}{P(\oplus)}$$

$$= \frac{(.01)(.98)}{.0395} = \frac{98}{395} \doteq 0.25$$

DG ch. 3

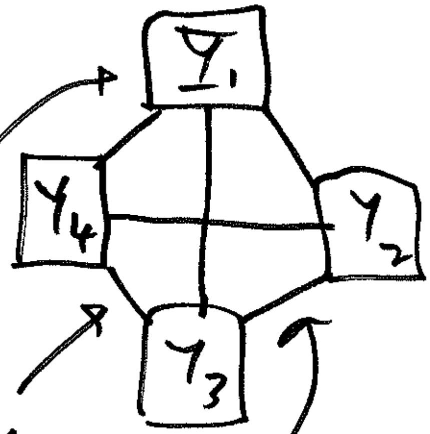
Random variables

$$P(Y = y)$$

random variable (process) (potential)

a possible value of Y (outcome) (kinetic)

(graph theory)



nodes

edge

$\boxed{Y_i \text{ dependent}}$

4 nodes

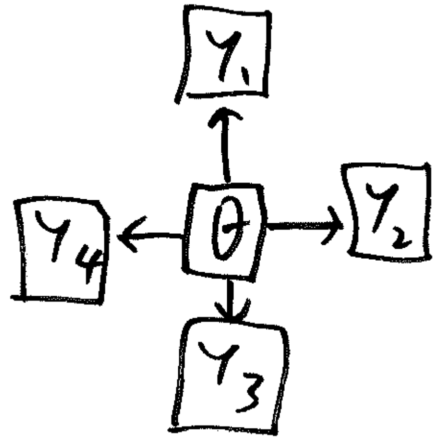
6 edges

x — y

x & y are dependent

$x \rightarrow y$

x causes y



5 nodes

4 edges

n nodes ← # data points

$\binom{n}{2}$ edges $= \dfrac{n(n-1)}{2}$

$= \underline{O}(n^2)$

this is of order n^2

= Big "oh" of n^2

$(n+1)$ nodes

n edges $= O(n)$

now Y_i are conditionally independent given θ

$O(n^2)$ / in this model, the past and the future are conditionally independent given the truth (θ)