07/29/19

Lecture 1 AMS 131

Tay-sachs case study
- noncarrier : 100% of normal Hex A
- carrier : 50% of normal Hex A (one H gene, one h)
- T-S : 0% of normal Hex A

n = # of children
as n↑, P(one or more T-S) ↑
P(A): The probability of A
'A' can be a set or a true/false (proposition)

probability: quantification of uncertainty
uncertainty: state of incomplete information

1650: Pascal + Fermat - first introduction to probability
↳ study of gambling (8000 years ago vs. 350 years ago)

Classical theory of probability
ex: amount of deer on campus
(cronic wasting disease)

subjects (n or N) $\begin{bmatrix} \text{datasets} \end{bmatrix}$ variables (p or K)

$N \doteq 850$
↑
approximately

$$\begin{bmatrix} Y \\ N \\ N \\ \vdots \\ Y \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \leftarrow \text{quantitative}$$

disease? qualitative

(AT RANDOM)
taking sample from population $n \begin{bmatrix} \ \end{bmatrix}$ population mean $\theta = p = ?$ uncertainty!

To decrease your uncertainty, get more <u>good</u> data.

As n↑, uncertainty about θ↓

good sample = representative of population
representative : similar to , in all relevant ways.

AMS 131 Lecture 1 (cont.)

Q: How to achieve representativeness?

A: Take sample at random!

with replacement	without replacement
IID: independent identically distributed	SRS: simple random sampling

SRS is more informative (don't want to double count), but IID is easier to calculate

$n = 1$ IID = SRS

$n = N$ no uncertainty about θ

$n << N$ IID $\approx$ SRS the probability of repeating is very low (ex: $\underset{N}{200{,}000{,}000}$ vs $\underset{n}{1{,}000}$)

$$\begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \xrightarrow[\text{(SRS)}]{\text{IID}} \begin{bmatrix} Y_1 \end{bmatrix} \quad n = 1 \qquad P(\text{odd}) = \frac{2}{3} = \frac{\text{"ways" favorable}}{\text{total "ways"}}$$

Equally-likely model: if all the number of ways an experiment can come out can be enumerated so that there is no reason one is favored over another $\longleftrightarrow$ elemental outcomes (EOs)

$$P(A) = \frac{\#\text{ of EOs that are favorable}}{\#\text{ of EOs}}$$

ex: the 1 and 7 are equally likely because IID is at random

Lecture 1 (cont.) AMS 131

Back to the T-S case

possible # of T-S kids

0
1
2
3
4
5

$P(1+ \text{ T-S kid}) = \frac{5}{6}$

if A, then B the equally likely model does not apply

If ELM, then $P(A) = \frac{5}{6}$

~~ELM~~, then $P(A) \neq \frac{5}{6}$

not all outcomes are equally likely

3 Types of Probability

1. Classical (Pascal-Fermat) ← probabilities are not weighted the same, ELM doesn't apply
2. Frequentist (1850) (John Venn)
3. Bayesian (1750) (Rev. Thomas Bayes)

Frequentist: relative frequency, repeatable under identical conditions independent of others - the probability P(A) is the long-run relative frequency" where A would occur in repetitions

ex: roullette

$P(\text{red}) = \frac{18}{38}$ ELM? yes!

0 green
00

1 red
2 black
3 red
...
36 black

spin #	outcome	red cumulative %
1	R	100%
2	~~G~~ not red	50%
3	R	67%
4	~~B~~ NR	50%

↓

converges to a limit $= \frac{18}{38}$

probability is defined by relative frequency.

Scanned by CamScanner

AMS 131 Lecture 1 (cont.)

Bayesian: 'A' can be any true/false proposition (not restricted to repeatability) and P(A) is a numerical measure of the weight of evidence in favor of the statement that 'A' is true.

Back to the T-S case (again):

$A = \{$ 1 or more T-S babies in a family of 5 children of 2 parents, both of whom are carriers (Hh) $\}$

- imagine a population of families because it's not repeatable in real life (families can't have 5 kids over and over)

		F: H	F: h
M	H	HH	Hh
M	h	Hh	hh

HH: 100% of normal of Hex A

Hh: 50% of normal (carrier)

hh: 0% of normal (T-S)

- no reason H or h should be favored (ELM)

$P(\text{normal}) = \frac{1}{4}$ $P(\text{carrier}) = \frac{2}{4}$ $P(\text{T-S}) = \frac{1}{4}$

One or more is logically equivalent to:

EXACTLY 1 OR 2 OR 3 OR 4 OR 5 T-S babies

Opposite of having:

EXACTLY 0 T-S babies = NOT { ↑ }

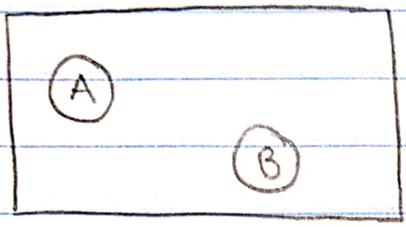
P(A or B) = P(A) ? P(B) P(A) vs. P(not A)

P(A and B) = P(A) ? P(B)

If looking at no T-S babies, the first must NOT be T-S, the second NOT T-S, third NOT T-S...

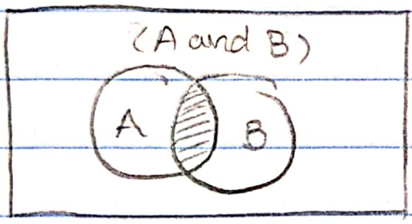
- connected by ANDs

Lecture 1 (cont.) AMS (3)



all possibilities
1. dart always hits box
2. once inside, each point is equally likely

$P(A \text{ or } B) = P(A) + P(B)$



$P(A \text{ or } B) = P(A) + P(B)$
- you double count the overlap,
intersection = AND

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$P(A \text{ or } B) = P(A) + P(B)$ special case: no overlap
when A and B are mutually exclusive

For any 'A', [0%] $0 \leq P(A) \leq 1$ [100%]
$P(A) = 0 \leftrightarrow$ A is false if and only
$P(A) = 1 \leftrightarrow$ A is true if.