4/9/19 Lecture 3

New on the class website: MSI tutoring times
Notetaker notes
All section times
Class syllabus
Quiz 2 and a new date for Quiz 1

The syllabus has a breakdown of what readings go with the topics discussed per lecture.

The first question is posted for Take-Home Test 1. More questions will be given as more topics are taught. David Draper will have an extra office hour the week the test is due.

## Case Study 2: AIDS

"Several studies of sexual partners of people infected w/ the virus shows that a single act of unprotected vaginal intercourse has a surprisingly low risk of infecting the uninfected partner — perhaps 1 in 100 to 1 in 1,000.

For an average, consider the risk to be 1 in 500. If there are 100 acts of intercourse w/ an infected partner, the chance of infection increases to 1 in 5.

Statistically, 500 acts of intercourse with 1 infected partner or 100 acts w/ 5 diff. infected partners lead to a 100% prob. of infection (statistically, not necessarily in reality)."

Evidently Dr. Shram would say that 600 acts would lead to a 120% prob. of infection (statistically speaking)

P(1 or more infected in 100 acts)

Side note: Don't go online for answers for hw and tests. The only place you should get help is from TAs and the prof.

$P(1 \text{ or more inf. in } 100 \text{ acts}) = 1 - P(0 \text{ inf. in } 100)$

$= 1 - P\left(\begin{smallmatrix}\text{not inf}\\\text{on 1st}\end{smallmatrix} \cap \begin{smallmatrix}\text{not inf}\\\text{on 2nd}\end{smallmatrix} \cap \dots \cap \begin{smallmatrix}\text{not inf}\\\text{on 100th}\end{smallmatrix}\right)$

These events are independent of each other due to the biology of the problem. (IID)

$= 1 - P\left(\begin{smallmatrix}\text{not inf.}\\\text{on 1st.}\end{smallmatrix}\right) \cdot P\left(\begin{smallmatrix}\text{not inf.}\\\text{on 2nd}\end{smallmatrix}\right) \cdots \cdot P\left(\begin{smallmatrix}\text{not inf.}\\\text{on 100th}\end{smallmatrix}\right)$

$= 1 - \left(1 - \frac{1}{500}\right)\left(1 - \frac{1}{500}\right) \cdots \left(1 - \frac{1}{500}\right)$ 1st order

$= 1 - \left(1 - \frac{1}{500}\right)^{100} = 0.18 = 18\%$

Dr. Schram said the prob. was $1/5 = 20\%$ which is not the same as $18\%$

$n = \#$ of acts $\quad p = P(\text{inf. on nth act})$

$P(\text{inf. in } n \text{ acts}) = 1 - (1-p)^n$

Q: Do you think that 500 acts w/ one infected partner is the same as 100 acts with 5 diff. infected partners?

A: No because people have different chances of infecting other people.

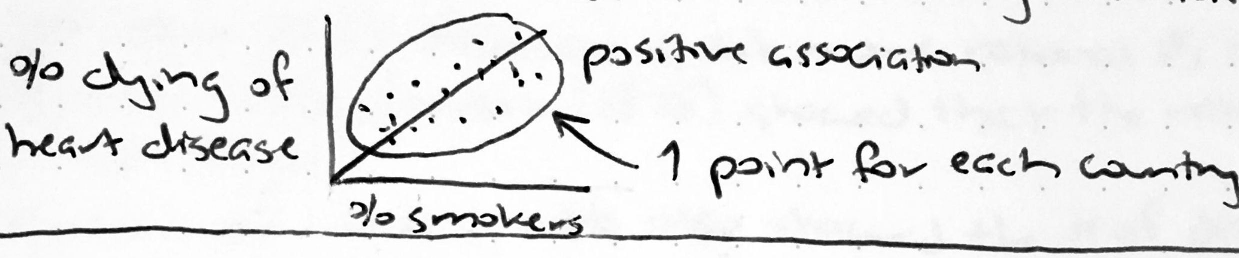
$P_i = P(\text{you get inf. from partner } i)$

$P(\text{inf in } n \text{ acts with } n \text{ partners}) = 1 - (1-p_1)(1-p_2)\dots(1-p_n)$

$= 1 - \prod_{i=1}^{n} (1-p_i)$

Ronald Aylmer Fisher advanced the "constitutional hypothesis": there is some genetic factor that disposes you both to smoke and die. Epidemiologists identified sets of twins where one smokes and the other one doesn't.

Sir Richard Doll (epidemiologist) found a strong association




Fisher's epidemiologists set up a race: which twin dies first — the smoker or the non-smoker?

$P(\text{both smokers die first}) = P(HH) = P(HH \mid H^*)$

$H^* =$ Fisher's theory

$= P(H \text{ on 1st and } H \text{ on 2nd}) = P(H \text{ on 1st}) \cdot P(H \text{ on 2nd})$

$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

2nd

| | | H | T |
|---|---|---|---|
| 1st | H | HH | HT |
| | T | TH | TT |

insufficient evidence to rule on $H^*$

$P(\text{all 9 smokers die first}) = \frac{1}{2^9} = \frac{1}{512} = 0.2\%$

Statistical Inference:

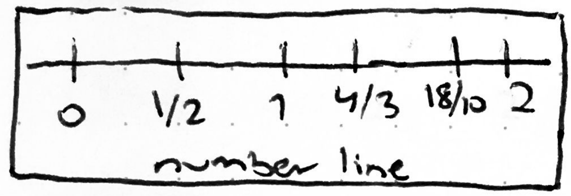
If $H^*$ is true, the data is highly unlikely. Therefore, $H^*$ is probably false.

This idea is a probabilistic version of proof by contradiction

Def: If the elements of an infinite set A can be placed in 1-to-1 correspondence with the positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$, A is said to be countably infinite.

Ex: The rational #s are real #s that can be expressed as ratios of integers (ex. $\frac{1}{2}, \frac{14}{13}, \frac{-89}{212}, \ldots$)

It might seem that there are a lot more rational #s than integers, but Georg Cantor (1878) showed that the rational #s are countable.




He also showed the # of distinct values on the real number line is an order of infinity greater than the # of integers or rationals.
"Some infinities are bigger than others."

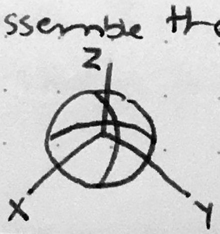
Def: An infinite set that is not countable is "uncountable"

Ex: $\mathbb{N} = \{1, 2, 3, \ldots\}$ is countable but $\mathbb{R} = \{$all real #s$\}$ is not countable

The mathematical foundation Kolmogorov chose for his development of probability theory is a part of mathematics called measure theory: an attempt to make rigorous the informal concepts of length, area and volume introduced by ancient Greek mathematics including Euclid and Pythagoras

However, people in the early 1900s discovered that $\infty$ is a weird thing when you try to make an idea like volume of a sphere in a 3-Dimensional space rigorous.

Thm: (Banach-Tarski paradox (1924)) Given a sphere in a 3-D space of radius 1, you can break up the sphere into a finite # of non-overlapping subsets (or pieces), move the pieces around, by rotating them and shifting them in the x, y, or z directions, and reassemble them into 2 identical copies of the original ball



Kolmogorov found that when $S$ is infinite, the set $2^S$ of all subsets is "too big" and "too strange" to permit the assignment of probabilities to all the sets in $2^S$ in a logically-internally-consistent way

When $S$ is infinite, Kolmogorov was forced to restrict attention to a smaller collection of subsets of $S$ than $2^S$ where nothing weird can happen.

The sets in this collection $\mathcal{C}$ have to satisfy 3 rules to avoid the weirdness

1. $\mathcal{C}$ includes the entire sample space $S$.

2. If an event $A$ is in $\mathcal{C}$ then so is its complement $A^c$.

(For 3) __Def__: Given any 2 sets $A$ and $B$, the union of $A$ and $B$ (written $A \cup B$ or $B \cup A$) is the set formed by throwing all the elements of $A$ and all the elements of $B$ together in one (potentially bigger) set (and discarding any and all duplicates)

This can be extended to more than 2 sets: if $A_1, A_2, \dots, A_n$ are events, we can talk about

$$A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^{n} A_i \quad \text{and}$$

if $A_1, A_2, \dots$ is a countable collection of events, we can even talk about

$$A_1 \cup A_2 \cup \dots = \bigcup_{i=1}^{\infty} A_i$$

3. If $A_1, A_2, \dots$ are all in $\mathcal{C}$ then so is $\bigcup_{i=1}^{\infty} A_i$

__Ex__: Whenever $|S| < \infty$ we can take $\mathcal{C} = 2^S$ with no weirdness arising, in other words if the sample space $S$ is finite, we can meaningfully assign probabilities to all of the subsets of $S$

Basic Facts about sets:

① For any event $A$, $(A^c)^c = A$

② $\emptyset^c = S$ and $S^c = \emptyset$

③ For any events $A, B$: $A \cup B = B \cup A$,

$A \cup A = A$, $A \cup A^c = S$, $A \cup \emptyset = A$,

$A \cup S = S$, and if $A \subset B$ then $A \cup B = B$

④ For any events $A, B, C$,
$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$
called associativity of the $\cup$ operation

<u>Def</u>: With any two sets $A$ and $B$, the intersection $A \cap B$ is the set of containing all, and only, those elements belonging both to $A$ and to $B$.

| Set operation | logical operation |
|---|---|
| $A^c$ | not $A$ |
| $A \cup B$ | $A$ or $B$ |
| $A \cap B$ | $A$ and $B$ |

(true/false propositions)

a set: a subset of $S$

If $A$ is an event, we can equivalently talk either about the set $A$ or the true/false proposition that one of the elements in $A$ is the outcome of the experiment $E$.

<u>Ex</u>: T-S disease $A = \{NNNNN\}$ as a set is equivalent to the T/F proposition (exactly 0 T-S babies) being true

⑤ Intersection of more than 2 sets:
$A_1, \dots, A_n \rightarrow A_1 \cap \dots \cap A_n = \bigcap_{i=1}^{n} A_i$

$A_1, A_2, \dots$ is $\bigcap_{i=1}^{\infty} A_i$

⑥ $A, B, C$ any events: $A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$
(Associativity of the $\cap$ operations)

**Def:** Two sets $A, B$ are disjoint = mutally exclusive if $A \cap B = \emptyset$ (if they have no outcomes in common

$n$ sets $A_1, \ldots, A_n$ are disjoint if all distinct pairs are disjoint: $A_i \cap A_j = \emptyset$ for $i \neq j$

Propositions $A, B$ being mutually exclusive corresponds to that they cannot both be true simultaneously

**Ex:** T-S disease $\begin{pmatrix}\text{Exactly} \\ \text{1 T-S} \\ \text{baby}\end{pmatrix}$ and $\begin{pmatrix}\text{Exactly} \\ \text{2 T-S} \\ \text{babies}\end{pmatrix}$ are mutually exclusive

(7) Any two sets $A, B$: $\left.\begin{aligned}(A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c\end{aligned}\right\}$ DeMorgan's Laws

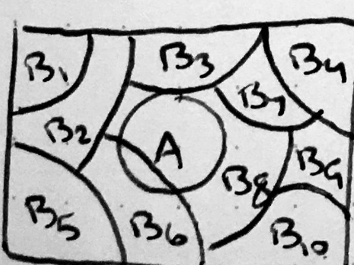
a) If $(A \cup B)^c$ is true, then $(A \cup B)$ is false, which can only occur if $A$ and $B$ are both false, making $A^c \cap B^c$ true.

b) if $(A \cap B)^c$ is true, then $A \cap B$ is false, which will occur if either one or both of $A, B$ are false, making $A^c \cup B^c$ true

(8) Any sets $A, B, C$: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$\begin{pmatrix}\text{distributive} \\ \text{property}\end{pmatrix}$ $\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(9) **Def:** If you can find events $B_1, \ldots, B_n$ such that
(a) the $B_i$ are mutually exclusive, and
(b) the $B_i$ are exhaustive in the sense that $\bigcup_{i=1}^{n} B_i = S$
then $(B_1, \ldots, B_n)$ forms a partition of $S$.




jigsaw puzzle $S$

$\begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix}$ – partition of $S$

$P(A) = P[(A \text{ and } B_1) \text{ or } (A \text{ and } B_2) \text{ or} \ldots \text{ or } (A \text{ and } B_n)]$

mut. exc. $= P(A \text{ and } B_1) + \ldots + P(A \text{ and } B_n)$

The idea of a partition is that every outcome in S lives inside one and only one of the partition sets

For any event $A$, $A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$
in other words,
$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

The partition chops A up into n mutually exclusive pieces (some of which may be empty) whose union is A

Kolmogorov's probability axioms

$P_K(A)$ needs to be a function from $\mathbb{C}$ (the collection of non-weird subsets of the real # line $\mathbb{R}$

Axiom 1) For all events $A \in \mathbb{C}$, $P_K(A) \geq 0$
(motivated by relative frequency)

Axiom 2) $P_K(S) = 1$ (motivated by relative frequency)

Axiom 3) For every countable collection of disjoint events $A_1, A_2, \ldots \in \mathbb{C}$,

$$P_K\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_K(A_i)$$ called countable additivity


if $A_1$ is approximately the same as $A_2$ ($A_1 \cong A_2$)
then $P_K(A_1) \cong P_K(A_2)$ ← continuity property

if $A_1, \ldots, A_n$ are disjoint then $P_K\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P_K(A_i)$