Lecture 6: Independence

Ex. COVID has a prevalence rate of 1%
We test and get + or −

↳ test accurately reports + 95% (sensitivity)

↳ // − 99% (specificity)

Q: I get + test,
what's the prob I have COVID?

D = have COVID) $P(D) = .01$
D^c = don't have COVID) $P(D^c) = .99$

$P(+ \mid D) = .95$; $P(- \mid D) = .05$
$P(- \mid D^c) = .99$; $P(+ \mid D^c) = .01$

Want: $P(D|+)$

$$= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)}$$

$$= \frac{(.95)(.01)}{(.95)(.01) + (.01)(.99)}$$

$$\approx .49$$

Independence:

Laymen's defn:

→ things don't affect each other

→ Knowing if A occurred or not, doesn't affect prob. of B.

Defn: We say A is independent of B denoted $A \perp B$ if

$$P(AB) = P(A)P(B).$$

→ distributive law for intersection

→ justifies intersection notation

<u>Theorem:</u> If $A \perp B$ then

$$P(A \mid B) = P(A).$$

Pf. $P(A \mid B) = \dfrac{P(AB)}{P(B)} = \dfrac{P(A)\cancel{P(B)}}{\cancel{P(B)}} = P(A).$

<u>Ex.</u> Roll two dice.

$P(\text{at least one } 6)$

$= 1 - P(\text{no } 6\text{s})$

$A_1 = \text{no } 6 \text{ on first roll}$

$A_2 =$ " second "

$= 1 - P(A_1 A_2)$ assume $A_1 \perp A_2$

$= 1 - P(A_1)P(A_2)$

$= 1 - (5/6)(5/6)$

$= 11/36$

<u>Counting perspective:</u>

Sampling $r=2$ from $n=6$.

w/ repl.

<u>Ordered:</u> $|S| = n^r = 36 = 6 \cdot 6$

$E =$ "at least one 6"

$= \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,6)$

$(6,1), (6,2), (6,3), (6,4), (6,5)\}$

$|E| = 11$

$$P(E) = \frac{|E|}{|S|} = 11/36.$$

<u>Unordered:</u>

$$|S| = \binom{n+r-1}{r} = \binom{6+2-1}{2} = \binom{7}{2} = 21$$

$E = \{\{1,6\},\{2,6\},\{3,6\},\{4,6\},\{5,6\},\{6,6\}\}$

$|E| = 6$

$$P(E) = \frac{|E|}{|S|} = \frac{6}{21} \neq 11/36.$$

Punchline: Ordered tends to give result assuming independence b/c of multiplicative structure.

Theorem:

If $A \perp B$ then

$\boxed{(1)\ A \perp B^c}$

(2) $A^c \perp B$

(3) $A^c \perp B^c$.

Pf. $P(AB^c)$
$= P(A) - P(AB)$
$= P(A) - P(A)P(B)$
$= P(A)(1 - P(B))$
$= P(A)P(B^c)$

Defn: Mutual Independence

Generalize to multiple events.

If (A_i) is a seq of events.

We say they are (mutually) independent

if for all subsequences

$$A_{i_1}, \dots, A_{i_k}$$

we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Q: Do I really need to check all subsequences? Yes.

Can I just check:

$$\boxed{P(A_1 A_2 A_3 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)?}$$

No.

Ex. Roll two dice (independently)

$A =$ "doubles" $= \{(1,1),(2,2),\ldots,(6,6)\}$

$|A| = 6$

$B =$ "Sum is between 7 and 10"

$= \{(1,6),(2,5),\ldots\}$

$|B| = 18$

$C =$ "Sum is 2, 7, 8"

$= \{(1,1),(1,6),(2,5),\ldots\}$

$|C| = 12$

Are these mutually independent?

$$P(ABC) \stackrel{?}{=} P(A)P(B)P(C)$$

$$1/36 = \left(\frac{6}{36}\right)\left(\frac{18}{36}\right)\left(\frac{12}{36}\right)$$

$$= \left(\frac{1}{6}\right)\left(\frac{1}{2}\right)\left(1/3\right) = \frac{1}{36}.$$

$$\overset{?}{P(BC) = P(B)P(C)} \qquad \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

→ "Sum is 7 or 8"

$= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(2,6),(3,5),(4,4),(5,3),(6,2)\}$

→ 11 outcomes.

$11/36$

So not mutually independent.

Can $A \perp A$?

$$P(A) = P(AA) = P(A)P(A) = P(A)^2$$

Works if $P(A) = 0$ or 1.

Ex. JWST had ~400 points of failure.

$$\boxed{T1} \to \boxed{T2} \to \boxed{T3} \to \cdots \to \boxed{T400}$$

JWST fails if any step fails.

What's prob. JWST works?

W_i = i^{th} task works

W_i^c = i^{th} task fails

Assume all tasks are independent.

$P(W_i^c) = 1/1000$

$P(\text{JWST works})$

$= P(W_1 W_2 W_3 \cdots W_{400})$

$= P(W_1)P(W_2)P(W_3) \cdots P(W_{400})$

$= \left(1-\frac{1}{1000}\right)\left(1-\frac{1}{1000}\right)\left(1-\frac{1}{1000}\right) \cdots \left(1-\frac{1}{1000}\right)$

$= \left(1-\frac{1}{1000}\right)^{400}$

$\approx .67$

↑ EXAM 1 ↑

Ex. Flip a coin 3 times.

$X = \#$ heads

$s \in S$	$X(s)$
H H H	3
H H T	2
H T H	2
H T T	1
T H H	2
T H T	1
T T H	1
T T T	0

Defn: Random Variable

A random variable (RV) X is a fn

$$X : S \to \mathbb{R}$$

also called - random variates,

- real-valued RVs
- univariate RVs

Ex. (1) toss two dice

$X = $ sum

(2) toss 25 coins, $X = $ length of longest run of Hs.