Nathanael Whitney
MTH4210
Professor Dyhr
May 10, 2019

Problem 1.15 DRAFT solution

1.15. Problem. (Section 3.4) A health study tracked a group of persons for five years. At the beginning of the study, 20% were classified as heavy smokers, 30% were classified as light smokers and the remaining were non-smokers. Results of the study showed that light smokers were twice as likely to die during the five-year study as non-smokers, but half as likely as heavy smokers. A randomly selected participant died over the five-year study. Find the probability that the participant was a heavy smoker.

Conditional probability:
We want to find: The probability that a randomly selected participant (1) died AND (2) was a heavy smoker.
Restated in conditional probability form:
"Given that a participant in the study died, find the probability that they were a heavy a smoker."

Restated in notation:
Let H = Heavy smoker
Let L = Light smoker
Let N = Non-smoker
Let D = Death
Find:
Prob(Death GIVEN Heavy Smoker) = Prob(Heavy Smoker | Death)
=> Find P(H | D)

Translate givens to probability statements:

Givens, part 1: (i) 20% were classified as heavy smokers (ii) 30% were classified as light smokers (iii) remaining were classified as non-smokers	**Translated:** (i) Prob(Heavy smoker) = 20% (ii) Prob(Light smoker) = 30% (iii) Prob(Non-smoker) = remaining
Convert to probability notation: P(H) = 0.20 P(L) = 0.30 P(N) = r	**Solve for r to get P(N):** P(N) = 1 – (P(H) + P(L)) = 1 – 0.20 + 0.30 = 1 – 0.50 = 0.50

Meaning we have the following probabilities:
P(H) = 0.20
P(L) = 0.30
P(N) = 0.50

Givens, part 2:
(iv) Light smokers were twice as likely to die during the study as non-smokers.
(v) Light smokers were half as likely to die as heavy smokers.

Restate in conditional probability form:
(iv) The probability that (1) will die AND (2) is a Light smoker.
"Given a participant is a Light Smoker, the probability of death is 2 times as likely as a Non-smoker."
Prob(Death | Light smoker) = 2 * Prob(Non-smoker)
P(D | L) = 2 * P(N)

(v) “Given a participant Death, the probability that they were a light smoker is one-half as likely as a Heavy smoker.”
Prob(Death | Light smoker) = 1/2 * Prob(Non-smoker)
P(D | L) = 1/2 * P(D | H)

Formula:
We will solve using Bayes’ Theorem, which informally is:
The probability of event B occurring, given event A occurring is:

$$\frac{\text{The probability of event B given event A occurring * The probability of A}}{\text{The sum of all the problem's conditional probability possibilities}}$$

Formally:

For j = 1, …, n

$$\mathbf{P(A_j|B) = \frac{P(BA_j)}{P(B)} = \frac{P(B|A_j) * P(A_j)}{P(B|A_j)*P(A_1) + P(B|A_2) + \ldots + P(B|A_j)*P(A_j) + \ldots + P(B|A_n)*P(A_n)}}$$

Translate from the formula to this problem:
$P(A_1)$ is P(H) = 0.20
$P(A_2)$ is P(L) = 0.30
$P(A_3)$ is P(N) = 0.50
P(B) is P(D) = unknown
P(D | H) = unknown
P(D | L) = 2 * P(D | N) → 1/2 * P(D | L) = P(D | N)
P(D | L) = 1/2 * P(D | H) → 2 * P(D | L) = P(D | H)
P(D | N) = unknown

Goal: Find P(D | H)

So we still have some unknowns. That most likely means we are hoping for some kind of insight or algebraic trick.

$$P(H \mid D) = \frac{P(D \mid H) * P(H)}{P(D \mid H) * P(H) + P(D \mid L) * P(L) + P(D \mid N) * P(N)}$$

Substitute probability knowns and see if that helps:

$$P(H \mid D) = \frac{P(D \mid H) * 0.20}{P(D \mid H) * 0.20 + P(D \mid L) * 0.30 + P(D \mid N) * 0.50}$$

Now, since we have those two algebraic equations, let’s try some algebraic substitutions.
2 * P(D | L) can be substituted for P (D | H) in the numerator and denominator:

$$P(H \mid D) = \frac{\mathbf{2 * P(D \mid L)} * 0.20}{\mathbf{2 * P(D \mid H)} * 0.20 + P(D \mid L) * 0.30 + P(D \mid N) * 0.50}$$

Similarly, 1/2 * P(D | L) can be substituted for P(D | N):

$$P(H \mid D) = \frac{2 * P(D \mid L) * 0.20}{2 * P(D \mid H) * 0.20 + P(D \mid L) * 0.30 + \mathbf{1/2 * P(D \mid L)} * 0.50}$$

Now everything is in terms of only one unknown, so we can do the algebra:
numerator: 2 * 0.20 = 0.40
denominator: (2 * 0.20) + 0.30 + (1/2 * 0.50) = 0.95

$P(D \mid H) = \frac{0.40\ \cancel{P(D \mid L)}}{0.95\ \cancel{P(D \mid L)}}$ and the unknown conditional probability cancels

Answer:

Therefore, $P(H \mid D) = \frac{0.40}{0.95} = 0.4210526 =$ $\boxed{\mathbf{0.42}}$