Raoul LePage
Professor
STATISTICS AND PROBABILITY
www.stt.msu.edu/~lepage
click on STT200_Sp09

3-18-09
GOLD STANDARD ANSWER

\{\$1a, \$1b, \$5c\}

Jack draws a bill first

Jill draws second

from the two bills then remaining

\[ P(\text{Jill } \$5) \]
\[ = \frac{2}{6} = \frac{1}{3} \]

same as Jack

all 6 possibilities

**Told Jack got \$1.**

\[ P(\text{Jill } \$5 | \text{Jack } \$1) = \frac{2}{4} \text{ Jill} \]

\[ \frac{2}{4} = \frac{2}{6} / \frac{4}{6} = \frac{P(\text{Jack } \$1 \text{ AND Jill } \$5)}{P(\text{Jack } \$1)} \]

**Define** \[ P(B | A) = \frac{P(AB)}{P(A)} \]
ARGUE WANT IN CLASSICAL SETUP

\[ P(B \mid A) = \frac{P(AB)}{P(A)} \]

GIVEN

DO NOT CARE IF A HAS HAPPENED

\[ \equiv \text{ MULTI RULE} \quad P(AB) = P(A)P(B \mid A) \]

= \[ P(B) = P(A \mid B) \]

CHOOSE THAT IMAGE HAVING \( P(\text{IMAGE} \mid \text{WHAT YOU SEE}) \) LARGEST
\[ P(A) = 0.3 \]
\[ P(B) = 0.4 \]
\[ P(AB) = 0.2 \]

**Total Probability Rule:**

\[ = P(AB) + P(AB') \]

**Solve!**

\[ P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.2}{0.3} = \frac{2}{3} \]

\[ P(B'|A) = 1 - P(B|A) = 1 - \frac{2}{3} = \frac{1}{3} \]

\[ P(B'|A') = 1 - P(B|A') = 1 - \frac{1}{7} = \frac{6}{7} \]

**Venn Diagram:**

- **A**
  - **AB**
    - **AB** = 0.2
    - **A'B** = 0.5
- **B**
  - **AB**
    - **AB** = 0.2
    - **A'B** = 0.2
- **BC**
  - **A'B'C** = 0.5

**Tree Diagram:**

- **P(A) = 0.3**
- **P(A') = 0.7**
- **P(B|A) = \frac{2}{3}**
- **P(B'|A) = \frac{1}{3}**
- **P(B'|A') = \frac{6}{7}**

**UC:**

- **P(A) → Venn → Tree → Venn**
\[ \text{JACK + JILL} \]
\[ P(JILL | JACK) = \frac{1}{2} \]
\[ \frac{1}{3} \quad \frac{1}{2} \]
\[ \frac{2}{3} \quad \frac{1}{2} = \frac{1}{3} \]
\[ \frac{2}{3} \quad \frac{1}{2} = \frac{1}{3} \]
\[ \frac{1}{3} \]
\[ 1 \]
“oil” = oil is present

“+” = a test for oil is positive

“−” = a test for oil is negative

false negative
false positive
\[
\begin{align*}
P(\text{oil}) &= 0.3 \\
P(+) | \text{oil} &= 0.9 \\
P(+) | \text{no oil} &= 0.4
\end{align*}
\]
TOTAL OF BRANCHES = 1

\[ P(\text{oil}) = 0.3 \]
\[ P(\text{+ | oil}) = 0.9 \]
\[ P(\text{+ | no oil}) = 0.4 \]
TOTAL OF CONDITIONAL BRANCHES = 1

\[ P(\text{oil}) = 0.3 \]

\[ P(+/\text{oil}) = 0.9 \quad P(-/\text{oil}) = 0.1 \]

\[ P(+/\text{no oil}) = 0.4 \]

sum of conditional probabilities is one
P(oil) = 0.3
P(+/oil) = 0.9
P(+/no oil) = 0.4
Oil contributes 0.27 to the total $P(+) = 0.55$. 
Oil contributes 0.27 of the total $P(+) = 0.27 + 0.28$. 

$P(\text{oil} | +) = P(\text{oil}+) / P(+) = 0.27 / (0.27 + 0.28) = 0.4909..$
The test for this infrequent disease seems to be reliable having only 3% false positives and 2% false negatives. What if we test positive?
We need to calculate $P(\text{diseased} \mid +)$, the conditional probability that we have this disease \textbf{GIVEN} we’ve tested positive for it.
CALCULATING OUR CHANCES OF HAVING THE DISEASE IF +

\[ P(+) = 0.0098 + 0.0297 = 0.0395 \]

\[ P(\text{disease} \mid +) = \frac{P(\text{disease} +)}{P(+) \times 100\%} = \frac{0.0098}{0.0395} = 0.248. \text{ only 25%!} \]
FALSE POSITIVE PARADOX
One may overwhelm a good test by failing to screen.

Even for this accurate test: P(diseased | +) is only around 25% because the non-diseased group is so predominant that most positives come from it.
FALSE POSITIVE PARADOX
one may overwhelm a good test by failing to screen

WHEN THE DISEASE IS TRULY RARE:
P(diseased | +) is a mere 3.2% because the huge non-diseased group has completely overwhelmed the test, which no longer has value
FOR MEDICAL PRACTICE: Good diagnostic tests will be of little use if the system is overwhelmed by lots of healthy people taking the test. Screen patients first.

FOR BUSINESS: Good sales people capably focus their efforts on likely buyers, leading to increased sales. They can be rendered ineffective by feeding them too many false leads, as with massive un-targeted sales promotions.