

Raoul LePage

Professor

STATISTICS AND PROBABILITY

www.stt.msu.edu/~lepage

click on STT200_Sp09

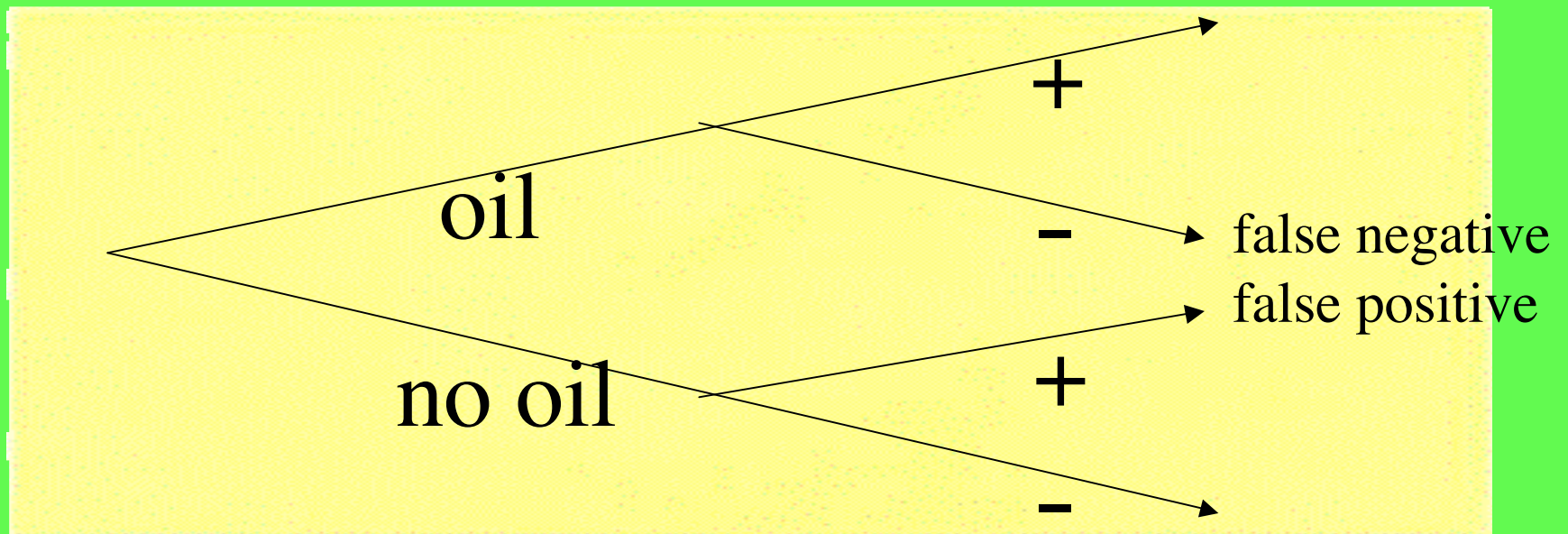
3-18-09

TREE DIAGRAM

“oil” = oil is present

“+” = a test for oil is positive

“-” = a test for oil is negative

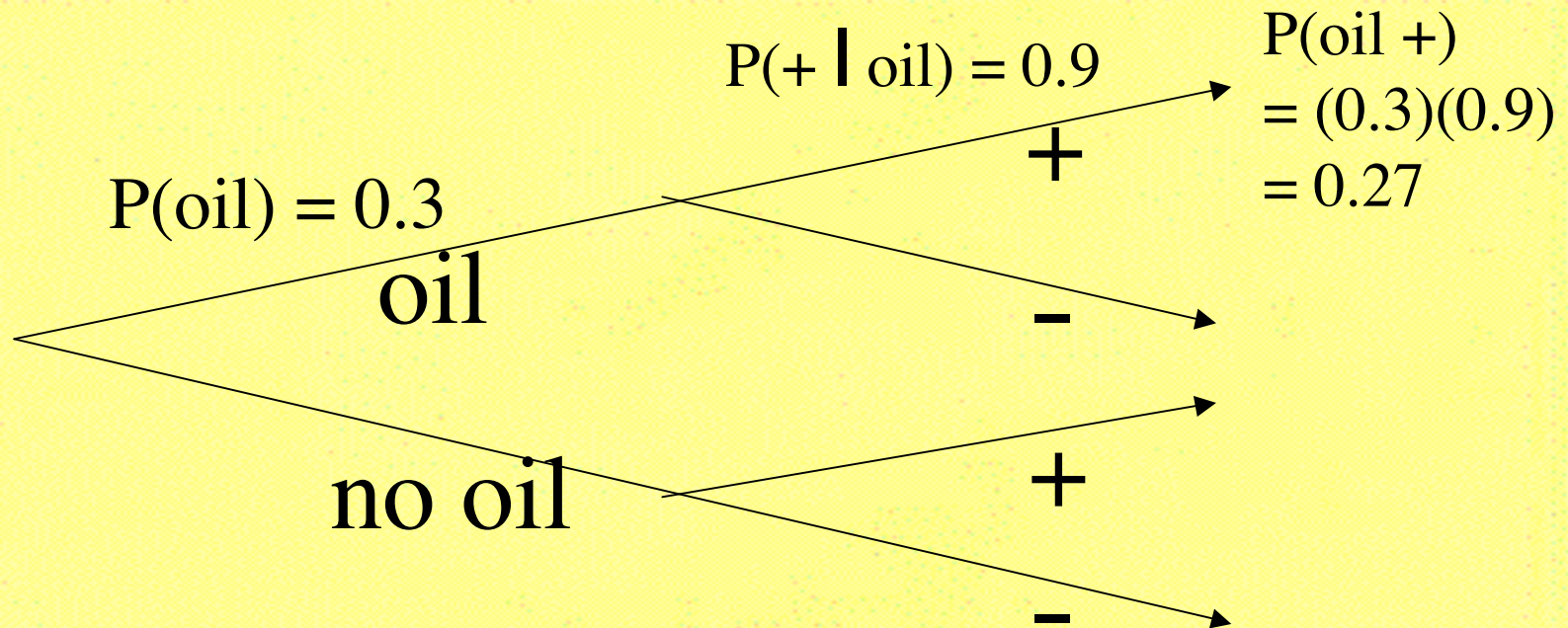


TREE DIAGRAM CONVENTIONS

$$P(\text{oil}) = 0.3$$

$$P(+ \mid \text{oil}) = 0.9$$

$$P(+ \mid \text{no oil}) = 0.4$$

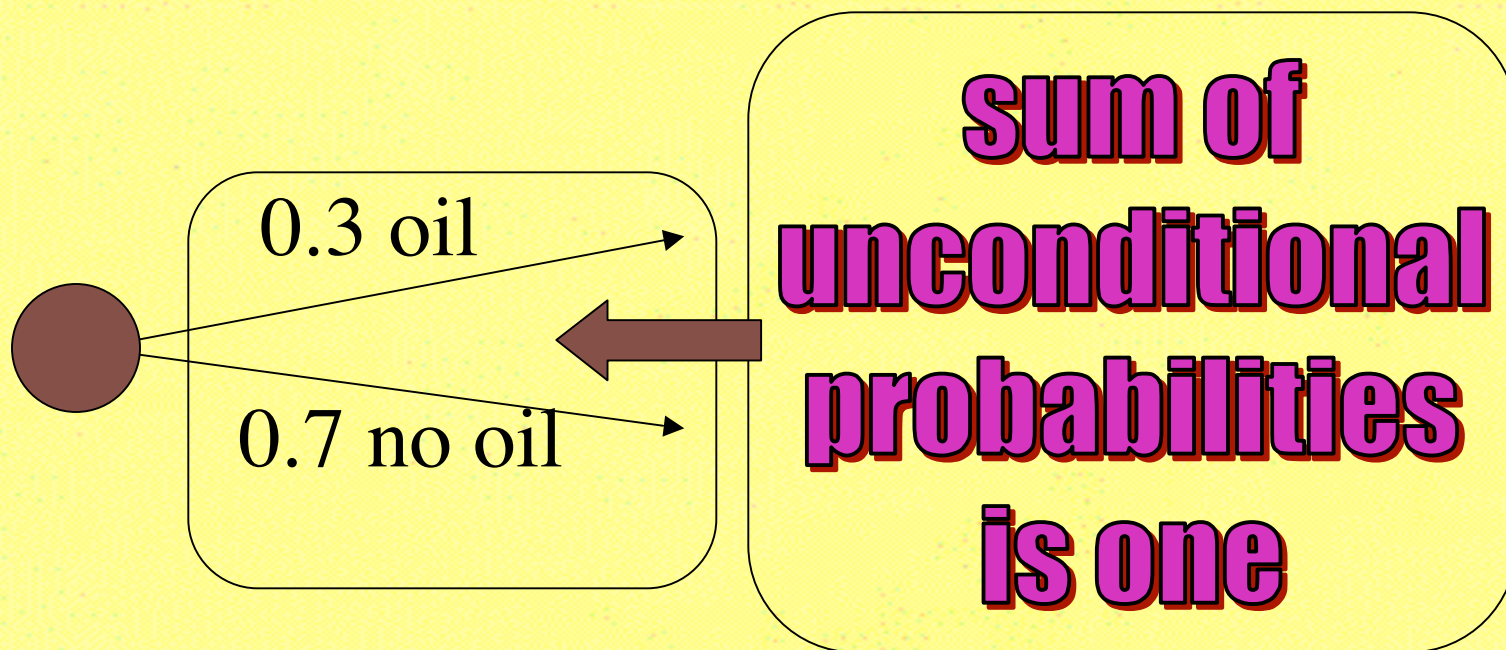


TOTAL OF BRANCHES = 1

$$P(\text{oil}) = 0.3$$

$$P(+ \mid \text{oil}) = 0.9$$

$$P(+ \mid \text{no oil}) = 0.4$$

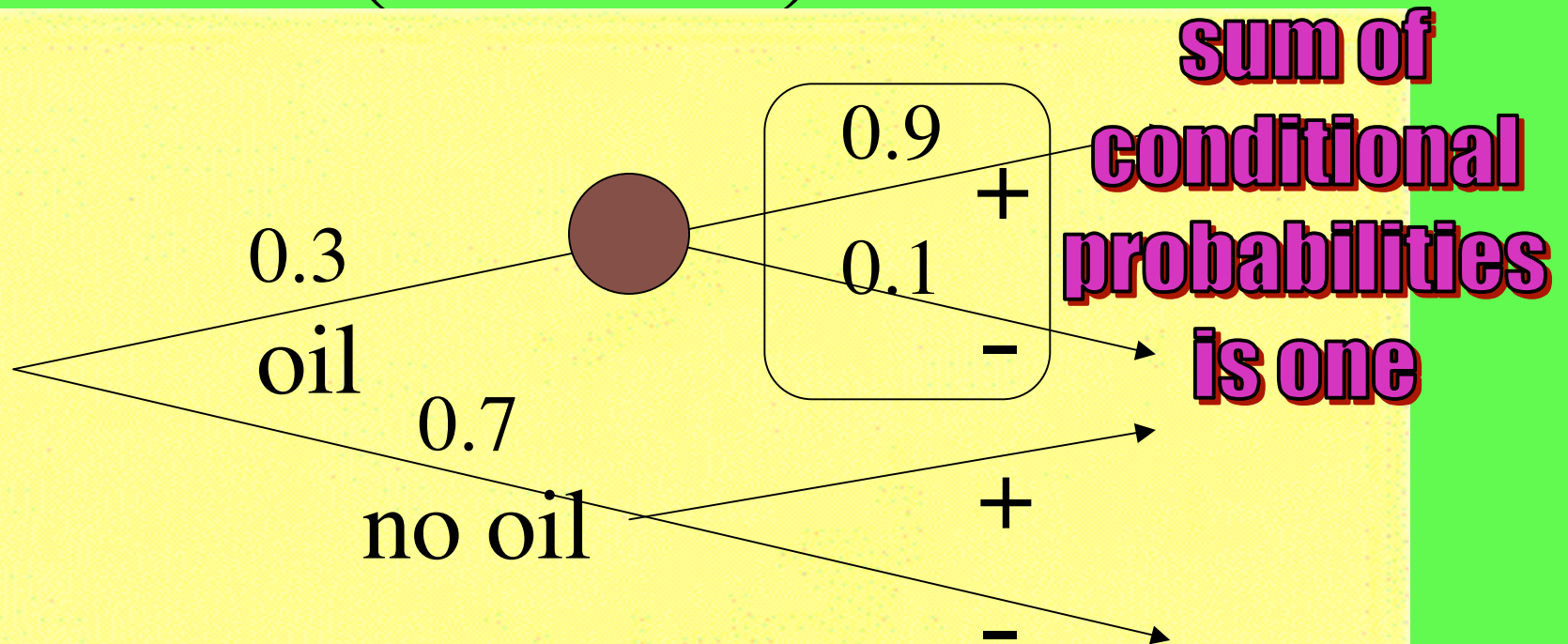


TOTAL OF CONDITIONAL BRANCHES = 1

$$P(\text{oil}) = 0.3$$

$$P(+ \mid \text{oil}) = 0.9 \quad P(- \mid \text{oil}) = 0.1$$

$$P(+ \mid \text{no oil}) = 0.4$$

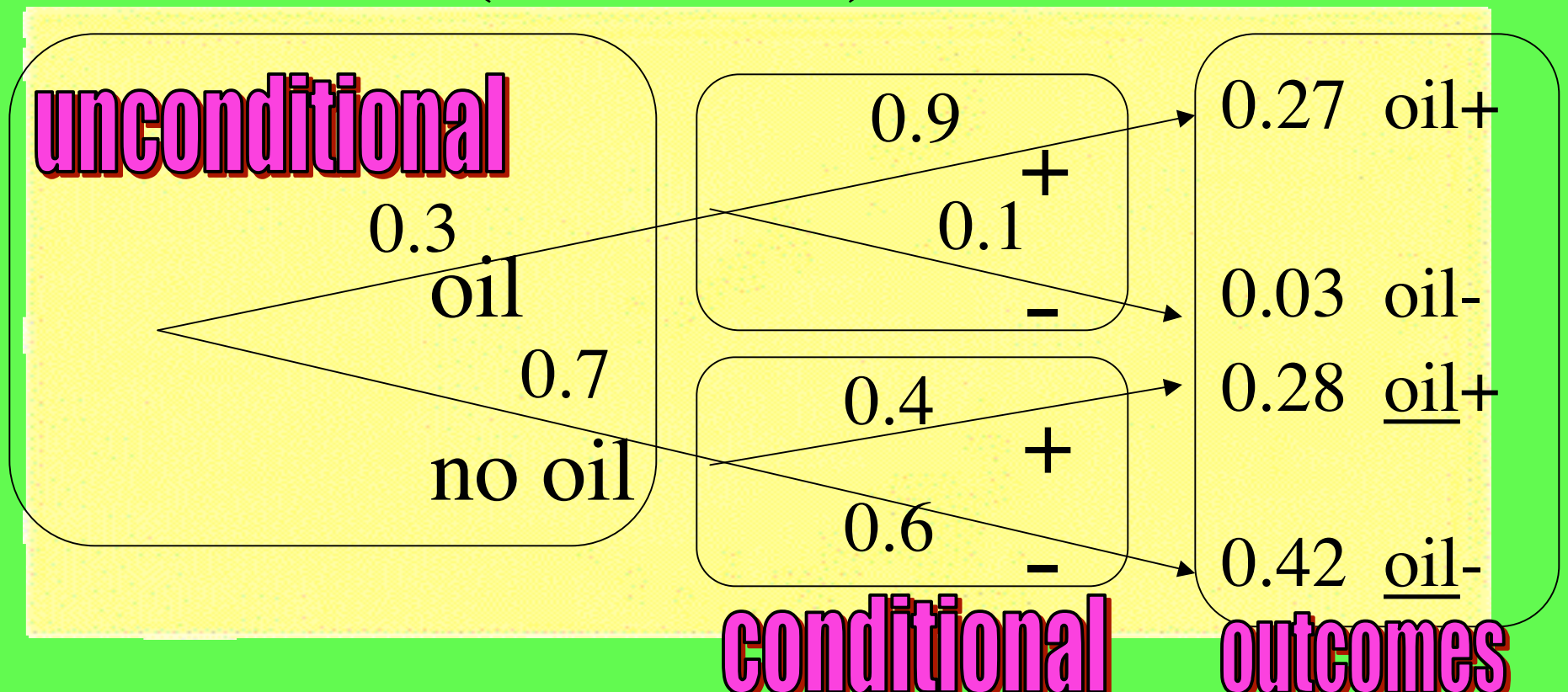


COMPLETE TREE

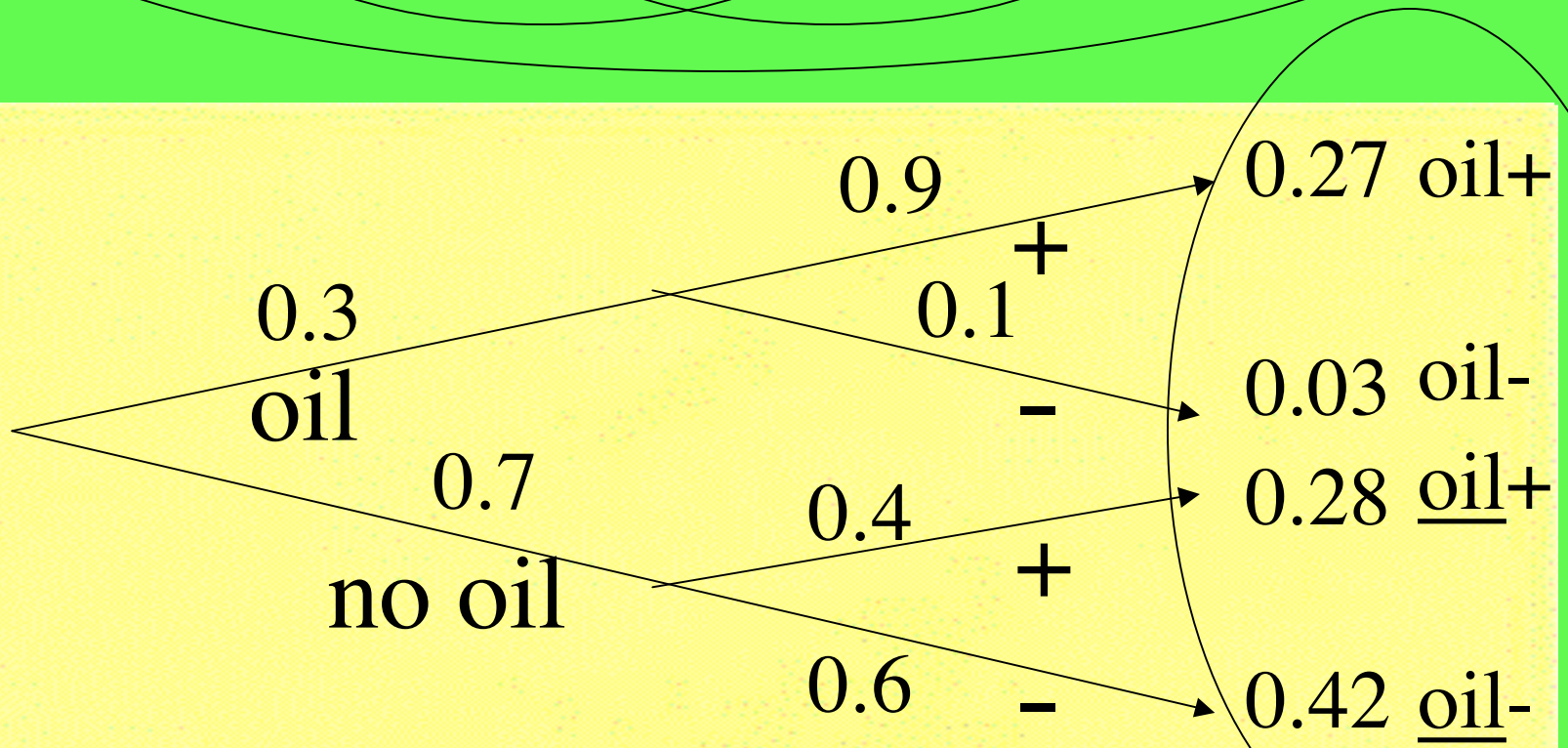
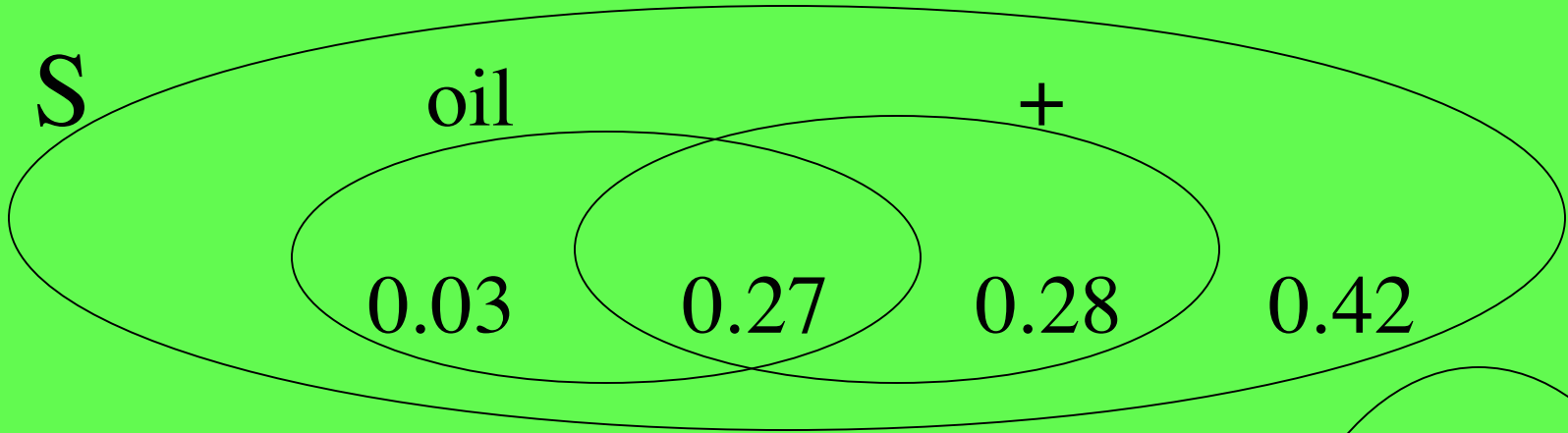
$$P(\text{oil}) = 0.3$$

$$P(+ \mid \text{oil}) = 0.9$$

$$P(+ \mid \text{no oil}) = 0.4$$



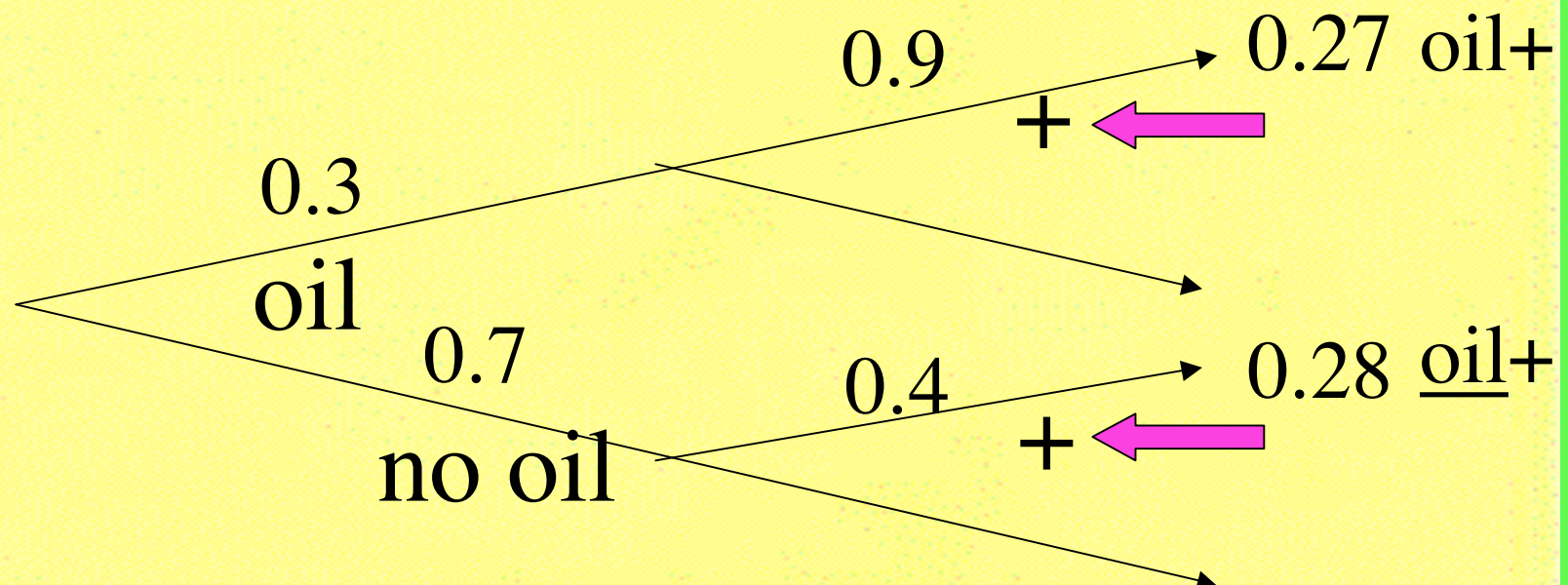
VENN DIAGRAM



TOTAL PROBABILITY

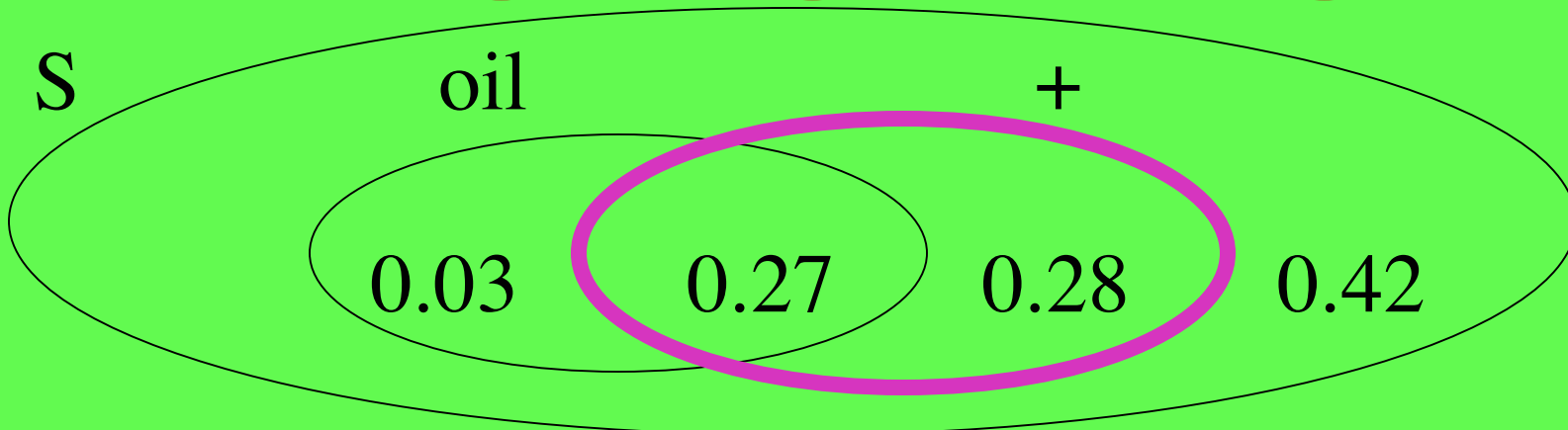
$$P(+)=P(\text{oil}+) + P(\text{no oil}+)$$

$$0.55 = 0.27 + 0.28$$



Oil contributes 0.27 to the total $P(+)=0.55$.

BAYES FORMULA



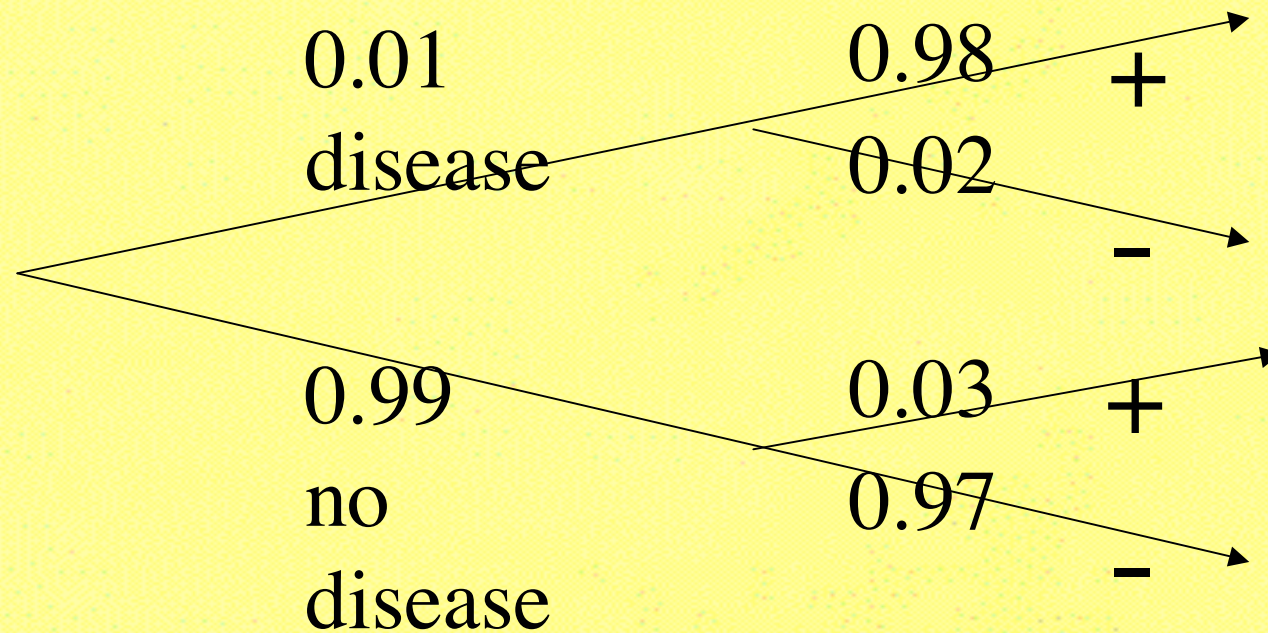
$$\begin{aligned} P(\text{oil} \mid +) &= P(\text{oil}+) / P(+) \\ &= 0.27 / (0.27 + 0.28) \\ &= 0.4909.. \end{aligned}$$

0.27 oil+

0.28 oil+

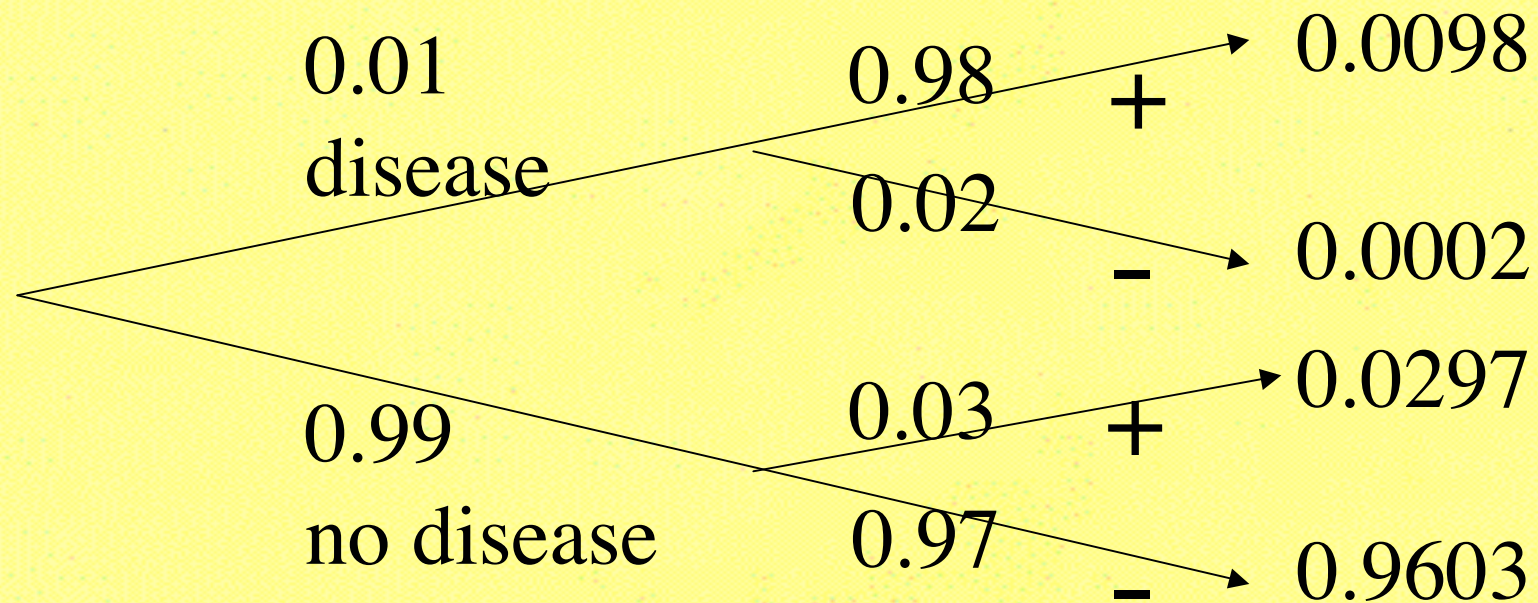
Oil contributes 0.27 of the total $P(+)$ = 0.27+0.28.

MEDICAL TEST



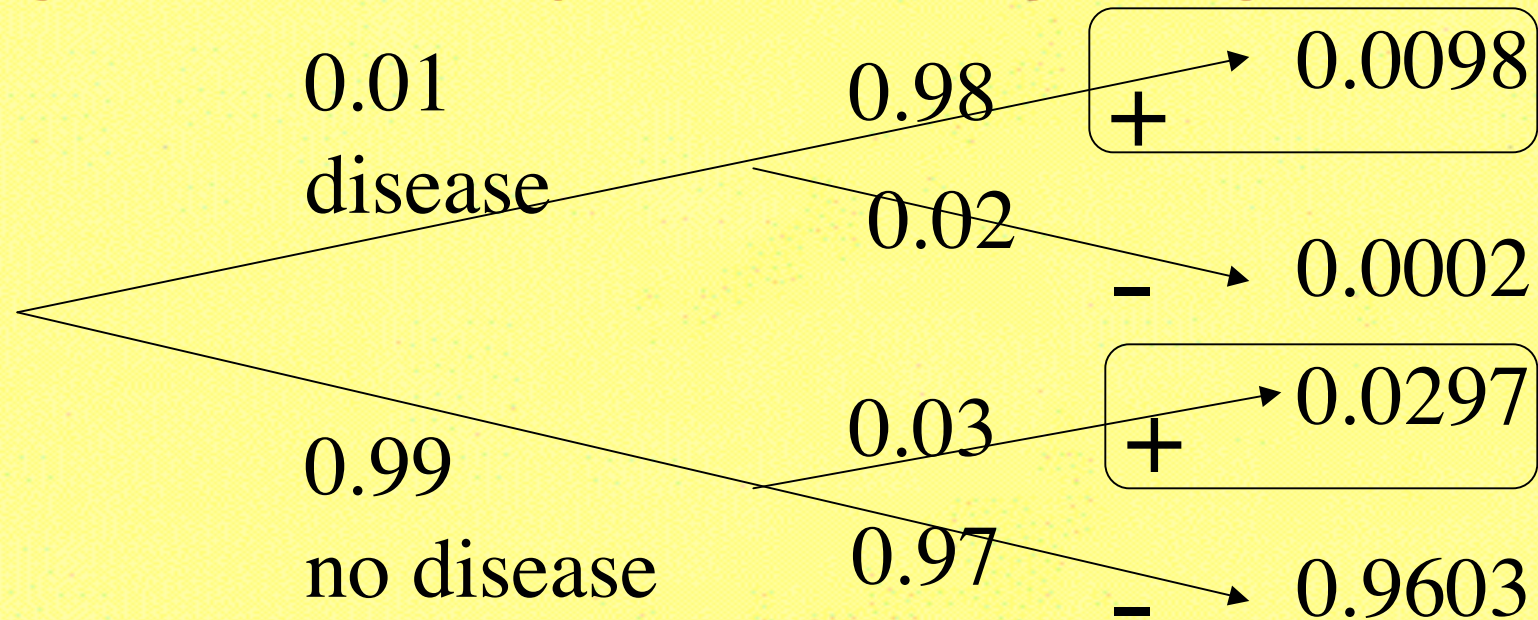
The test for this infrequent disease seems to be reliable having only 3% false positives and 2% false negatives. **What if we test positive?**

MEDICAL TEST



We need to calculate $P(\text{diseased} \mid +)$, the **conditional probability** that we have this disease **GIVEN** we've tested positive for it.

CALCULATING OUR CHANCES OF HAVING THE DISEASE IF +

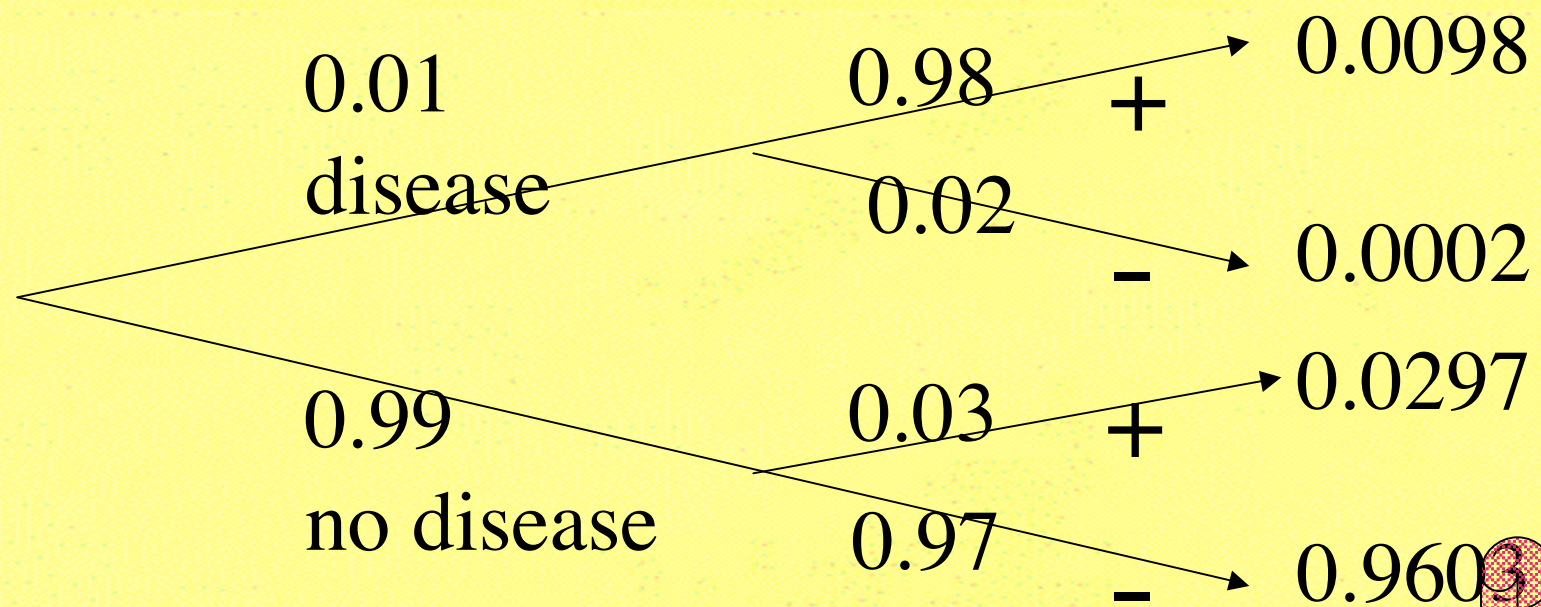


$$P(+) = 0.0098 + 0.0297 = 0.0395$$

$$P(\text{disease} \mid +) = P(\text{disease}+) / P(+)$$
$$= .0098 / 0.0395 = 0.248. \text{ only } 25\% !$$

FALSE POSITIVE PARADOX

one may overwhelm a good test by failing to screen

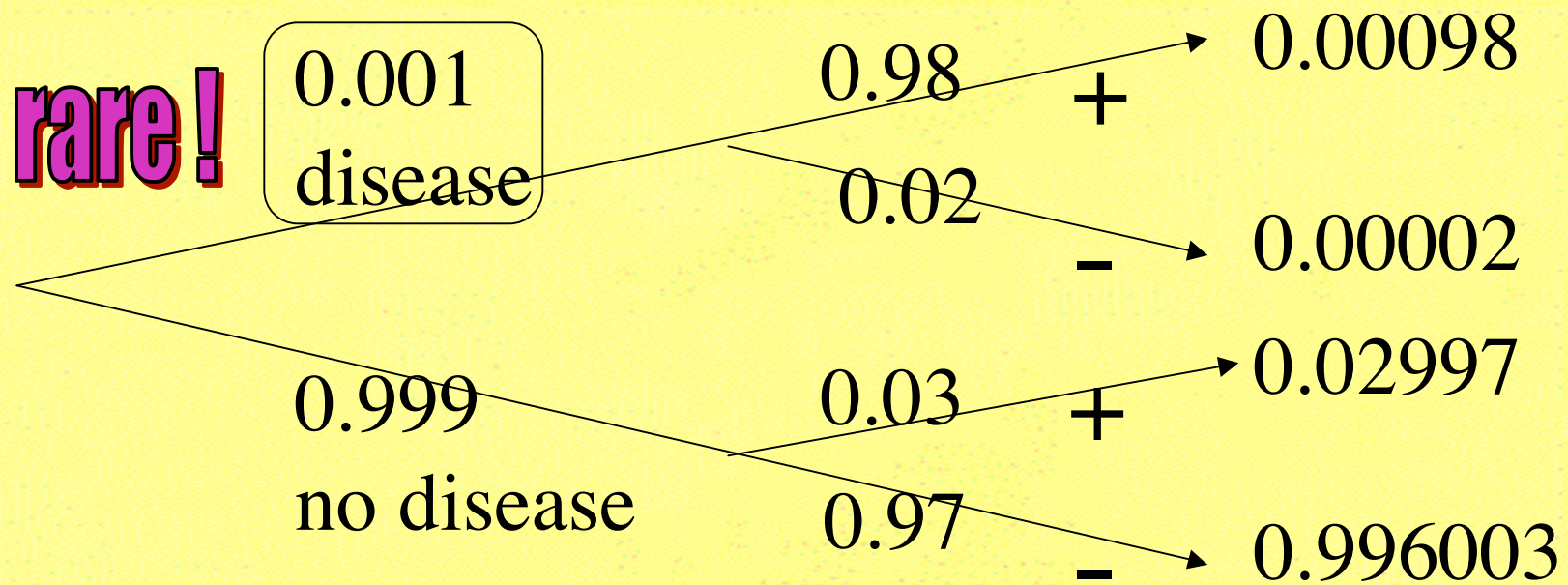


EVEN FOR THIS ACCURATE TEST:

$P(\text{diseased} \mid +)$ 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

IMPLICATIONS OF THE PARADOX

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.**

