Detecting spam email

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$ Bayes' Theorem

- 60% of all email in 2016 is spam.
- 20% of spam has the word "Dear"
- 1% of non-spam (aka ham) has the word "Dear"

$P(F) = 0.6$
$P(E|F) = 0.2$
$P(E|F^C) = 0.01$

You get an email with the word "Dear" in it.

What is the probability that the email is spam?

1. Define events & state goal
2. Identify known probabilities
3. Solve

Let: E: "Dear", F: spam
Want: $P(\text{spam}|\text{"Dear"}) = P(F|E)$

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)} = \frac{(0.2)(0.6)}{(0.2)(0.6) + (0.01)(0.4)}$$

≈ 0.967

Detecting spam email, an understanding

- 60% of all email in 2016 is spam.
- 20% of spam has the word "Dear"
- 1% of non-spam (aka ham) has the word "Dear"

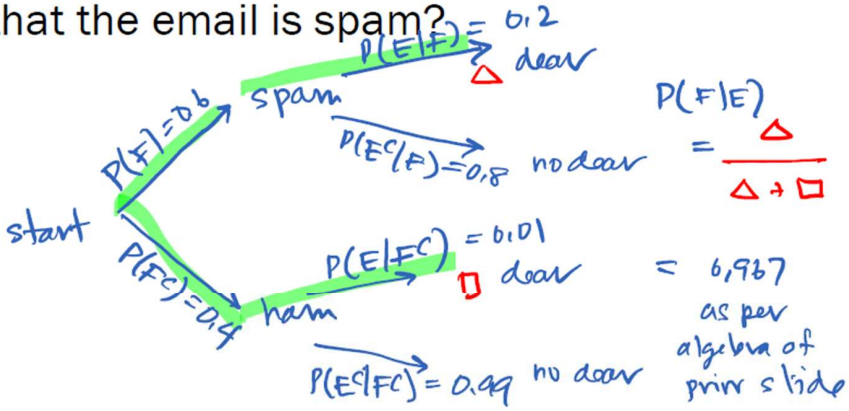
Note: You should know how to use Bayes/ Law of Total Prob., but drawing a tree can help.

You get an email with the word "Dear" in it.

What is the probability that the email is spam?

1. Define events & state goal

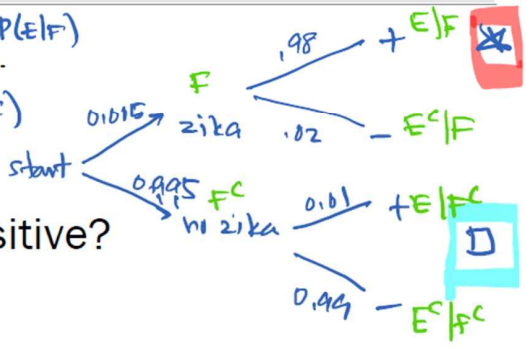
Let: E: "Dear", F: spam
Want: $P(\text{spam}|\text{"Dear"}) = P(F|E)$



Zika Testing

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$ Bayes' Theorem

- A test is 98% effective at detecting Zika ("true positive"). $P(E|F)$
- However, the test has a "false positive" rate of 1%. $P(E|F^C)$
- 0.5% of the US population has Zika. $P(F)$



What is the likelihood you have Zika if you test positive?

Why would you expect this number?

1. Define events & state goal
2. Identify known probabilities
3. Solve

Let: E = you test positive
F = you actually have the disease

Want:
P(disease | test+)
$= P(F|E)$

$$P(F|E) = \frac{(0.005)(0.98)}{(0.005)(0.98) + (0.995)(0.01)} \approx 0.330$$

Why it's still good to get tested

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$ Bayes' Theorem

- A test is 98% effective at detecting Zika ("true positive").
- However, the test has a "false positive" rate of 1%.
- 0.5% of the US population has Zika.

Let: E = you test positive
F = you actually have the disease

Let: E^C = you test negative for Zika with this test.

What is $P(F|E^C)$?

	F, disease +	F^C, disease −
E, Test +	True positive $P(E\|F) = 0.98$	False positive $P(E\|F^C) = 0.01$
E^C, Test −	False negative $P(E^C\|F) = 0.02$	True negative $P(E^C\|F^C) = 0.99$

$$P(F|E^C) = \frac{P(E^C|F)P(F)}{P(E^C|F)P(F) + P(E^C|F^C)P(F^C)} \approx 0.0061$$ via similar math