190F Foundations of Data Science

Spring 2020

Lecture 23

Decisions

Announcements

Decisions

The Medical Diagnosis Problem

- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don't.

- *Q1: What is the probability that a randomly selected person in the population has the disease?*
- *A1: $P(D) = 0.001$. This is called the prior probability of disease.*

The Medical Diagnosis Problem

- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don't.

- *Q2: What is the probability that a randomly selected person tests positive if they do have the disease?*
- *A2: $P(TP|D) = 0.99$. This is the likelihood of testing positive given that you have the disease.*

The Medical Diagnosis Problem

- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don't.

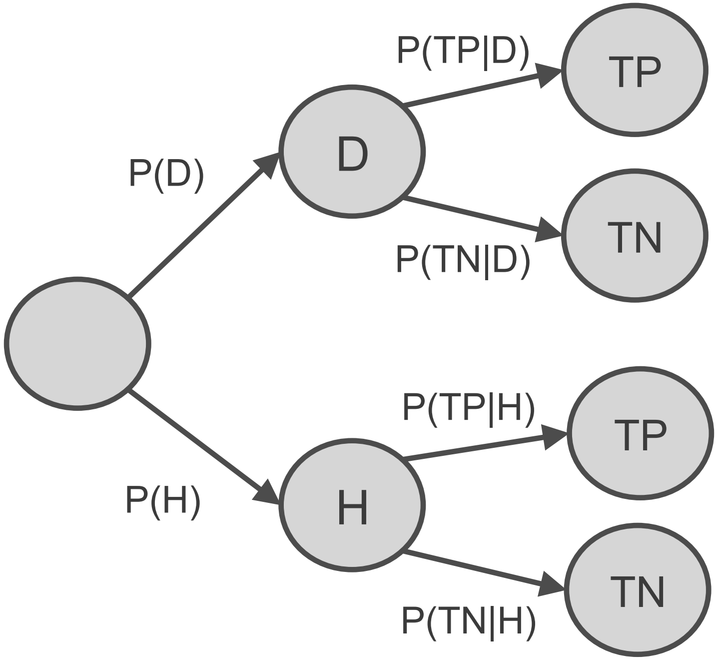
- *Q3: What is the probability that a randomly selected person tests positive if they do not have the disease (are healthy)?*
- *A3: P(TP|H)=0.05. This is the* false positive rate *of the test. It is the likelihood of testing positive given that you do not have the disease.*

The Medical Diagnosis Problem

- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don't.

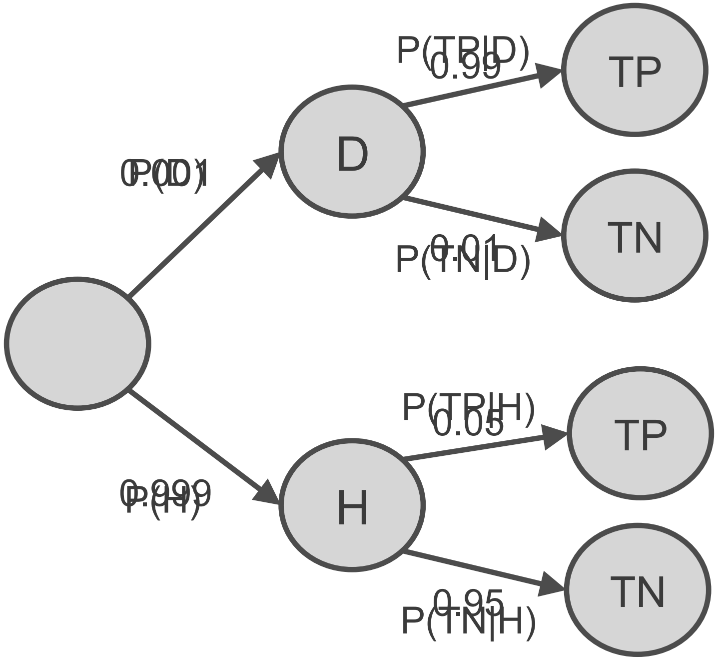
- *Q4: What is the probability that a randomly selected person is healthy and tests positive?*

Conditional Probability Tree



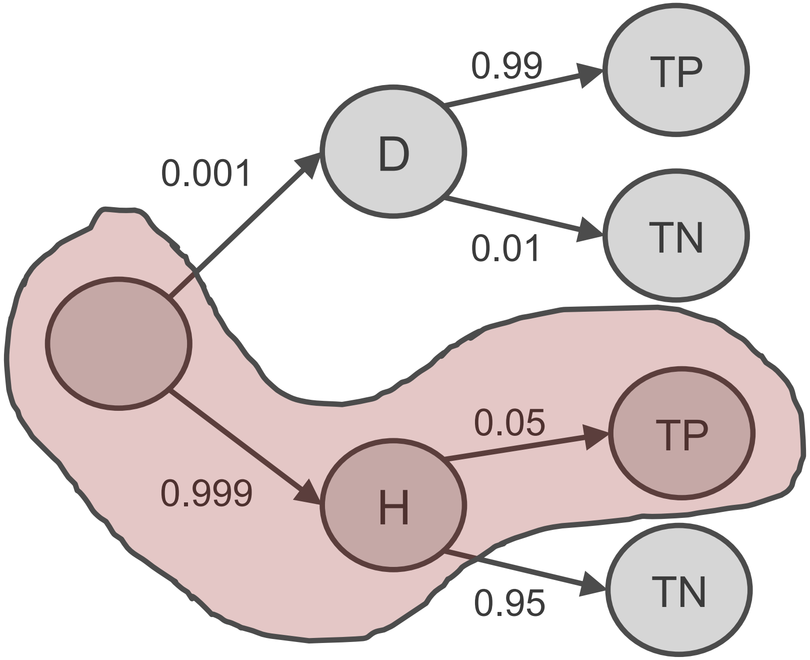
- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don’t.

Conditional Probability Tree



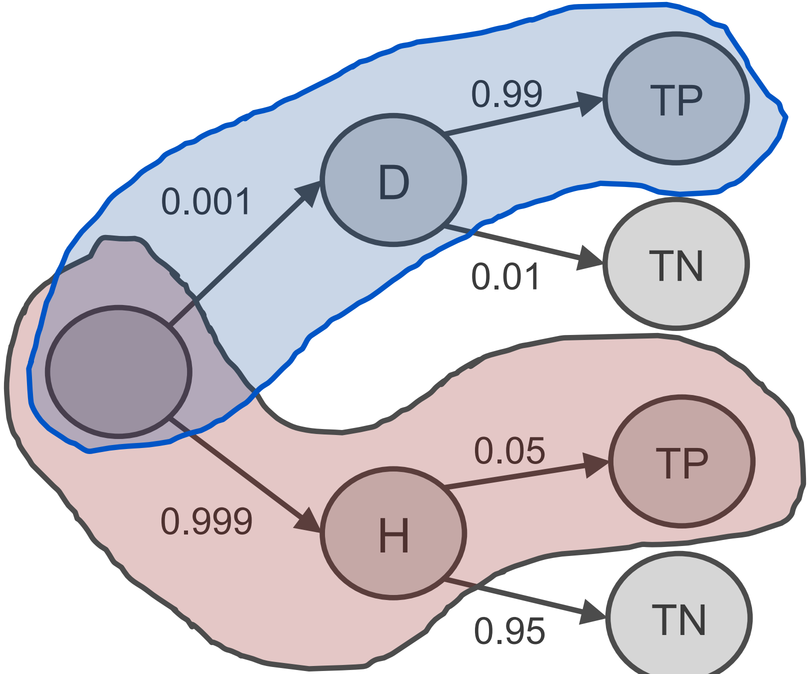
- Suppose there is a rare disease with a prevalence of 1/1000.
- Suppose that there is an imperfect test that indicates that you have the disease 99% of the time when you really do, and 5% of the time when you really don't.

Back to Questions…



- *Q4: What is the probability that a randomly selected person is healthy and tests positive?*
- *A4:* $P(H,TP)=P(H)P(TP|H)$
  $= 0.999 \times 0.05$
- *This is the joint probability of being healthy and testing positive.*

More Questions...

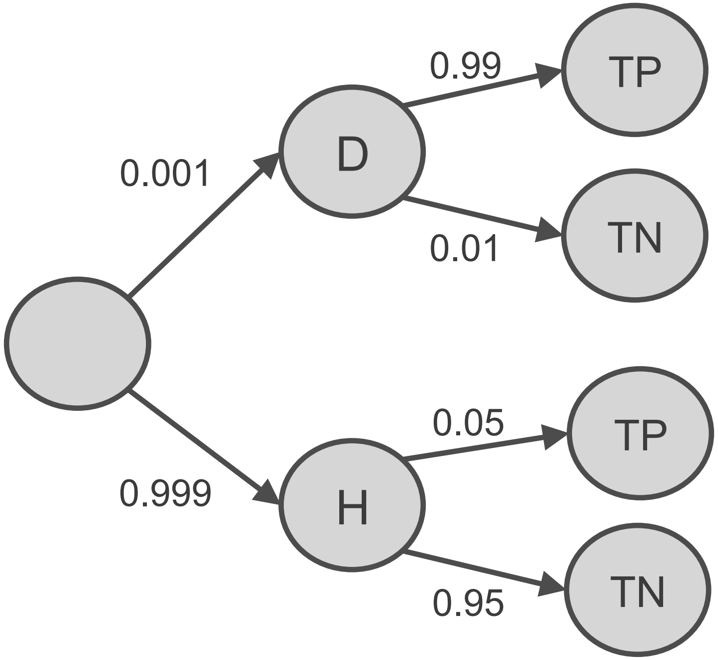


- *Q5: What is the probability that a randomly selected person tests positive?*
- *A5:* $P(TP)=P(H)P(TP|H) + P(D)P(TP|D)$

  $= 0.999 \times 0.05 + 0.001 \times 0.99$

  $= 0.05095$
- *This is the marginal probability of testing positive.*

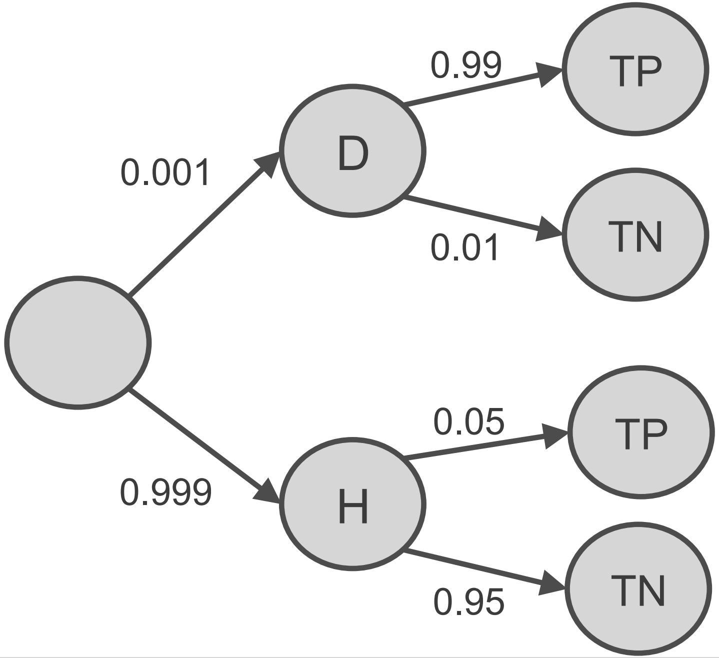
More Questions…



- *Q6: What is the probability that a randomly selected person has the disease if they test positive?*

Some Intuition…

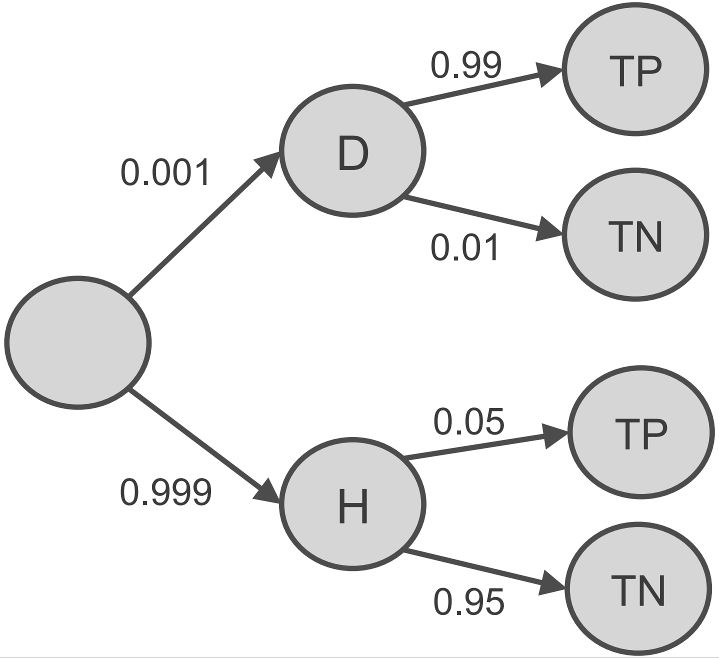
- Out of 1000 people, on average:



	Test Pos.	Test Neg.
Disease		
Healthy		

Some Intuition…

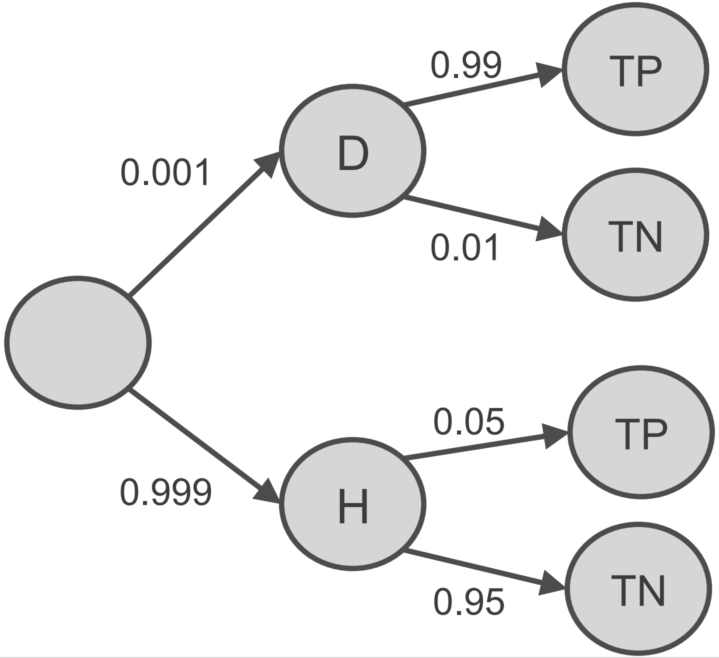
- Out of 1000 people, on average:



	Test Pos.	Test Neg.
Disease	~1	
Healthy		

Some Intuition…

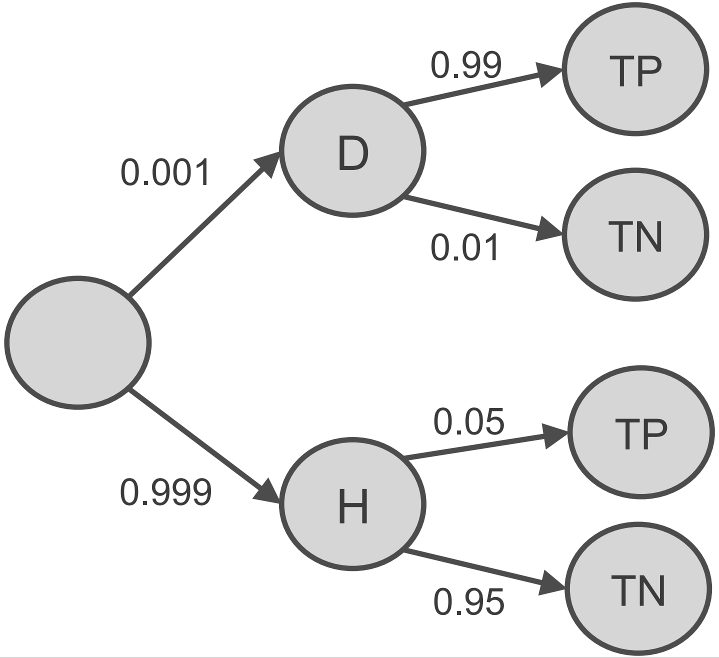
- Out of 1000 people, on average:



	Test Pos.	Test Neg.
Disease	~1	~0
Healthy		

Some Intuition…

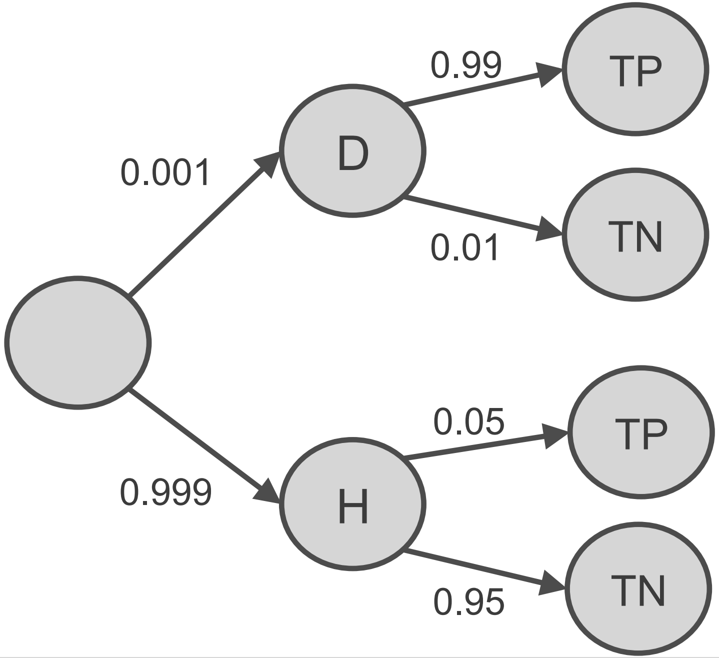
- Out of 1000 people, on average:



	Test Pos.	Test Neg.
Disease	~1	~0
Healthy	~50	

Some Intuition…

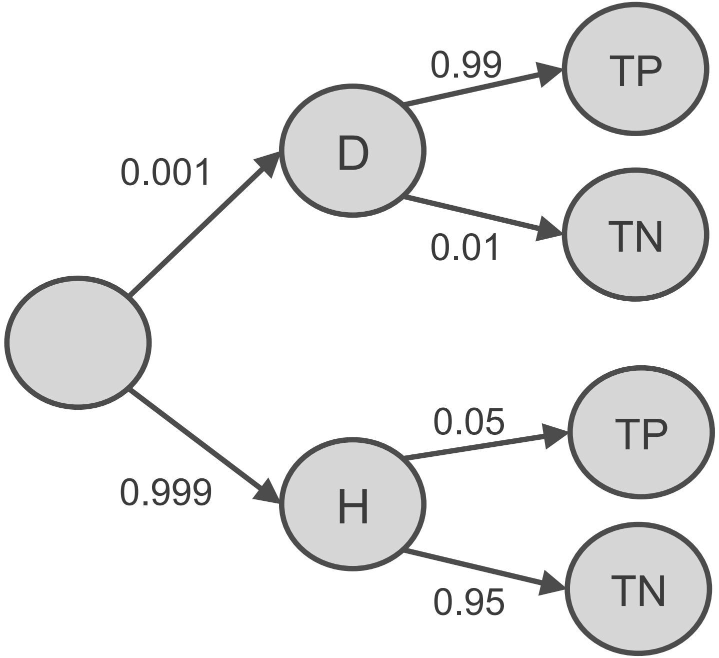
- Out of 1000 people, on average:



	Test Pos.	Test Neg.
Disease	~1	~0
Healthy	~50	~950

- So only ~1/50 or 2% of patients with positive test results have the disease.

The Math...



- $P(D|TP) = P(D,TP)/P(TP)$
- $P(D,TP) = P(TP|D)P(D)$
  $= 0.99 \times 0.001 = 0.00099$
- $P(TP) = P(D, TP) + P(H, TP)$
- $P(H,TP) = P(TP|H)P(H)$
  $= 0.05 \times 0.999$
  $= 0.04995$
- $P(D|TP) = 0.00099/0.0594 = 0.0194$

Bayes' Rule

$$P(D|TP) = \frac{P(TP,D)}{P(TP)}$$

Bayes' Rule

Posterior

Likelihood

Prior

$$P(D|TP) = \frac{P(TP|D)P(D)}{P(TP|D)P(D) + P(TP|H)P(H)}$$

$$P(D|TP) = \frac{P(TP,D)}{P(TP)}$$

The Medical Diagnosis Problem

"We asked 20 house officers, 20 fourth-year medical students and 20 attending physicians, selected in 67 consecutive hallway encounters at four Harvard Medical School teaching hospitals, the following question:

"If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?"

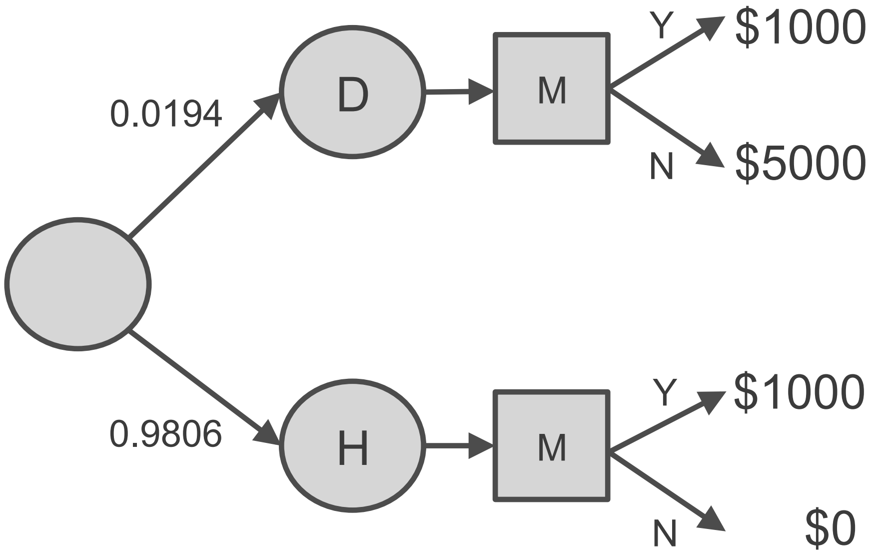
Casscells et al. Interpretation by Physicians of Clinical Laboratory Results. NEJM, 1978.

The Medical Diagnosis Problem

"Eleven of 60 participants, or 18%, gave the correct answer. These participants included four of 20 fourth-year students, three of 20 residents in internal medicine and four of 20 attending physicians. The most common answer, given by 27, was that [the chance that a person found to have a positive result actually has the disease] ***was 95%.***

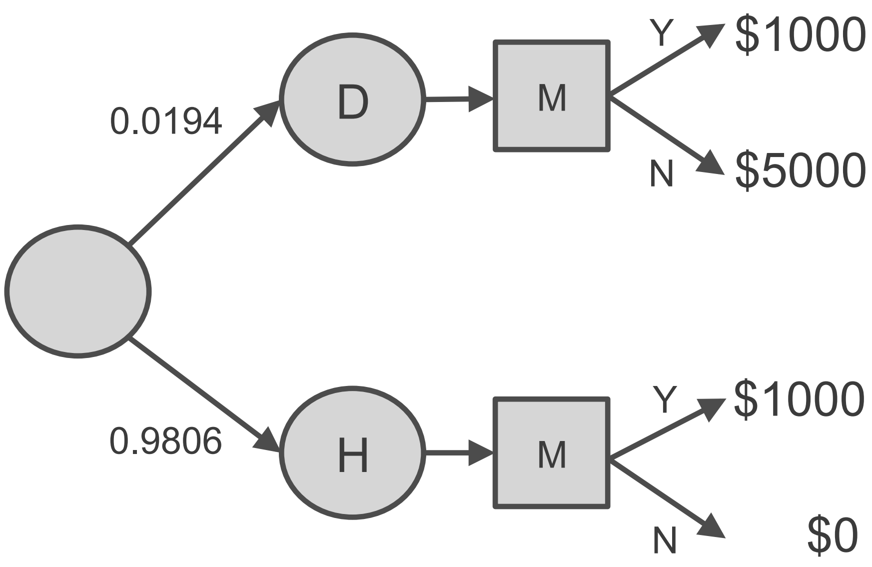
Casscells et al. Interpretation by Physicians of Clinical Laboratory Results. NEJM, 1978.

Treatment Costs



- Suppose a patient tests positive.
- There is an effective medication to treat the patient, but it costs $1000.
- If the patient has the disease and is not treated now, on average it will cost $5000 to treat them after the disease progresses.
- *Q6: Should we treat now?*

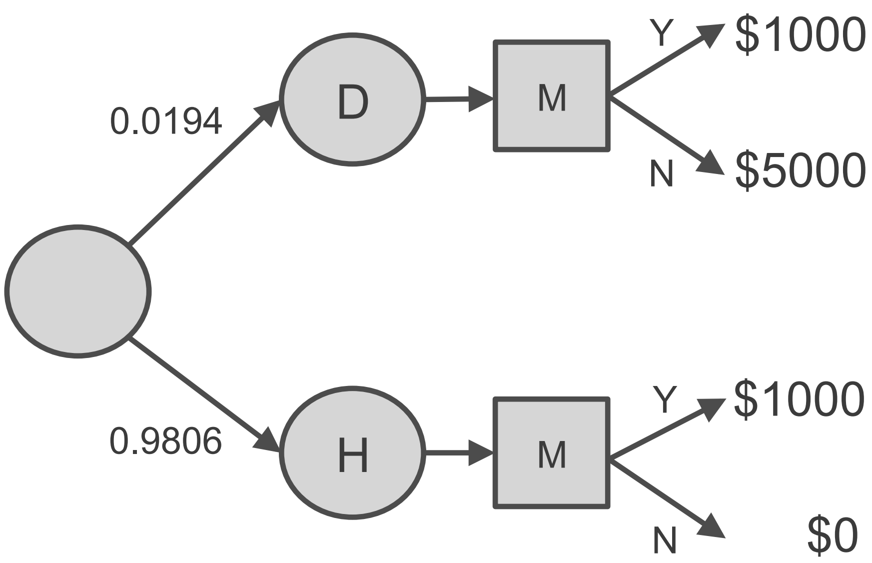
Treatment Costs



- What is the *expected cost* of treatment?
- $E[C|M=Y] = P(D)x1000 + P(H)x1000$

  $= \textit{\$1000}$

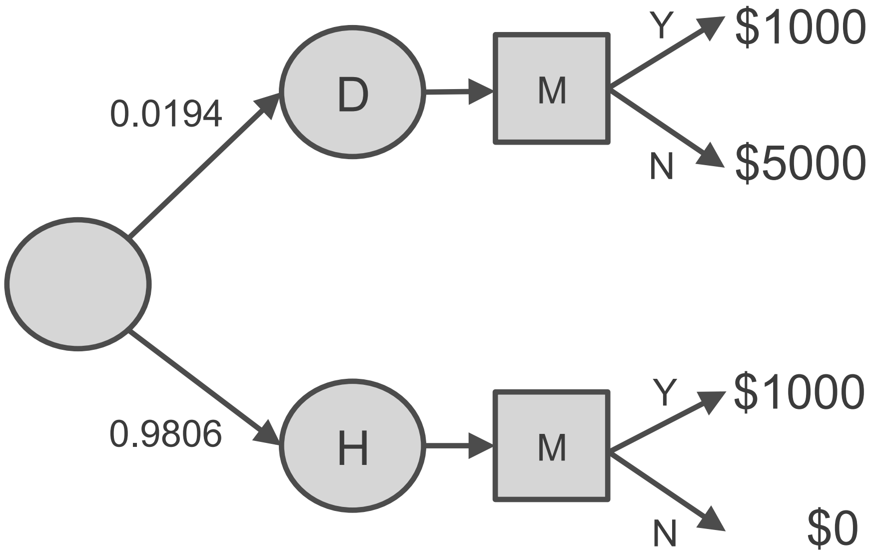
Treatment Costs



- What is the *expected cost* of not treating the patient?
- $E[C|M=N] = P(D)x5000 + P(H)x0$

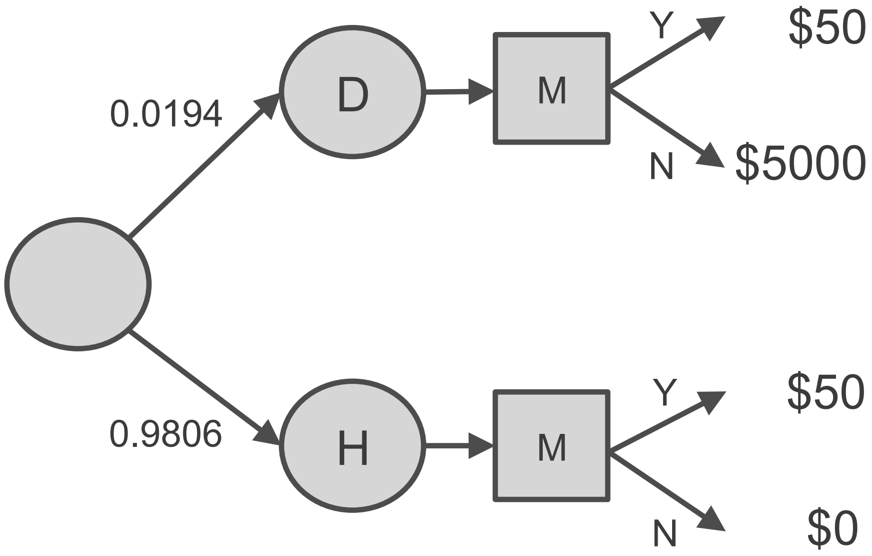
  $= \$97$

Treatment Costs



- *Q6: Should we treat now?*
- *A6: According to expected treatment cost, it costs less on average to not treat a patient with a positive test now. We should wait and treat the patient later if it turns out the really do have the disease.*

Effect of Medication Costs



- *Q7: How about now?*
- E[C|M=Y] = P(D)x50
  + P(H)x50
  = *$50*
- E[C|M=N] = P(D)x5000
  + P(H)x0
  = *$97*
- The expected costs support immediate treatment.