BAYES' THEOREM

Thomas Bayes (c. 1702 – 17 April 1761) was a British mathematician and Presbyterian minister, known for having formulated a specific case of the theorem that bears his name: Bayes' theorem, which was published posthumously.

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

The Problem

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

P(S) = ?

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

$P(S) = 0.2$

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

P(S) = 0.2
P(L given S) = ?

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

P(S) = 0.2
P(L given S) = 0.01

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

P(S) = 0.2
P(L given S) = 0.01
P(L given S^c) = ?

Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

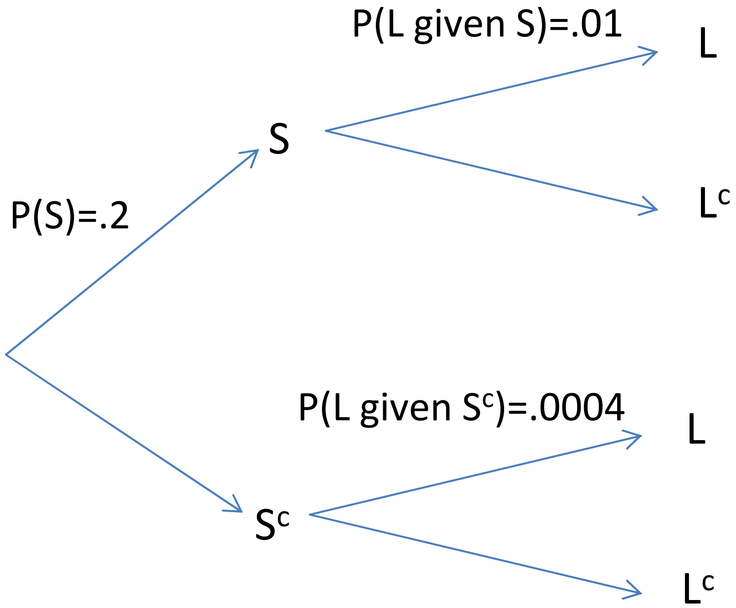
P(S) = 0.2
P(L given S) = 0.01
P(L given S^c) = 1/2500 = 0.0004

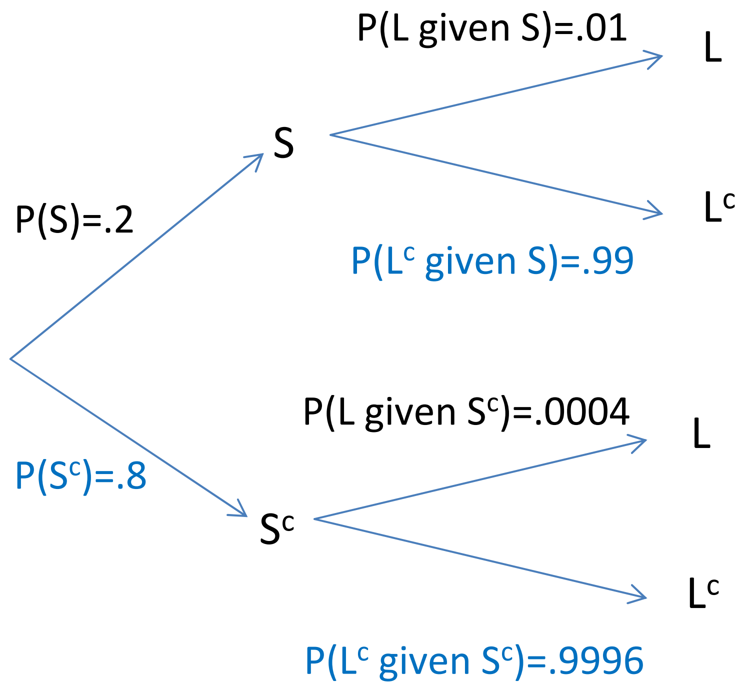
Approximately 20% of Americans smoke.
Of those who smoke, 1% develop lung cancer.
Of those who don't smoke,
1 in 2500 will get lung cancer.

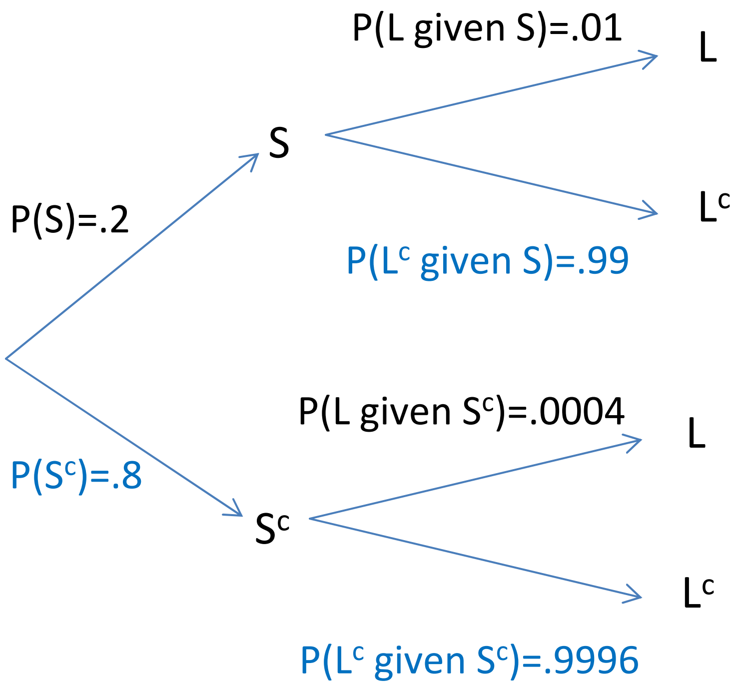
If someone has lung cancer, what is the probability that they smoke?

Let S = people who smoke
L = people with lung cancer

P(S) = 0.2
P(L given S) = 0.01
P(L given S^c) = 1/2500 = 0.0004
Find P(S given L)


P(L given S)=.01
L
S
L^c
P(S)=.2
P(L given S^c)=.0004
L
S^c
L^c


P(L given S)=.01
L
S
Lᶜ
P(S)=.2
P(Lᶜ given S)=.99
P(L given Sᶜ)=.0004
L
P(Sᶜ)=.8
Sᶜ
Lᶜ
P(Lᶜ given Sᶜ)=.9996

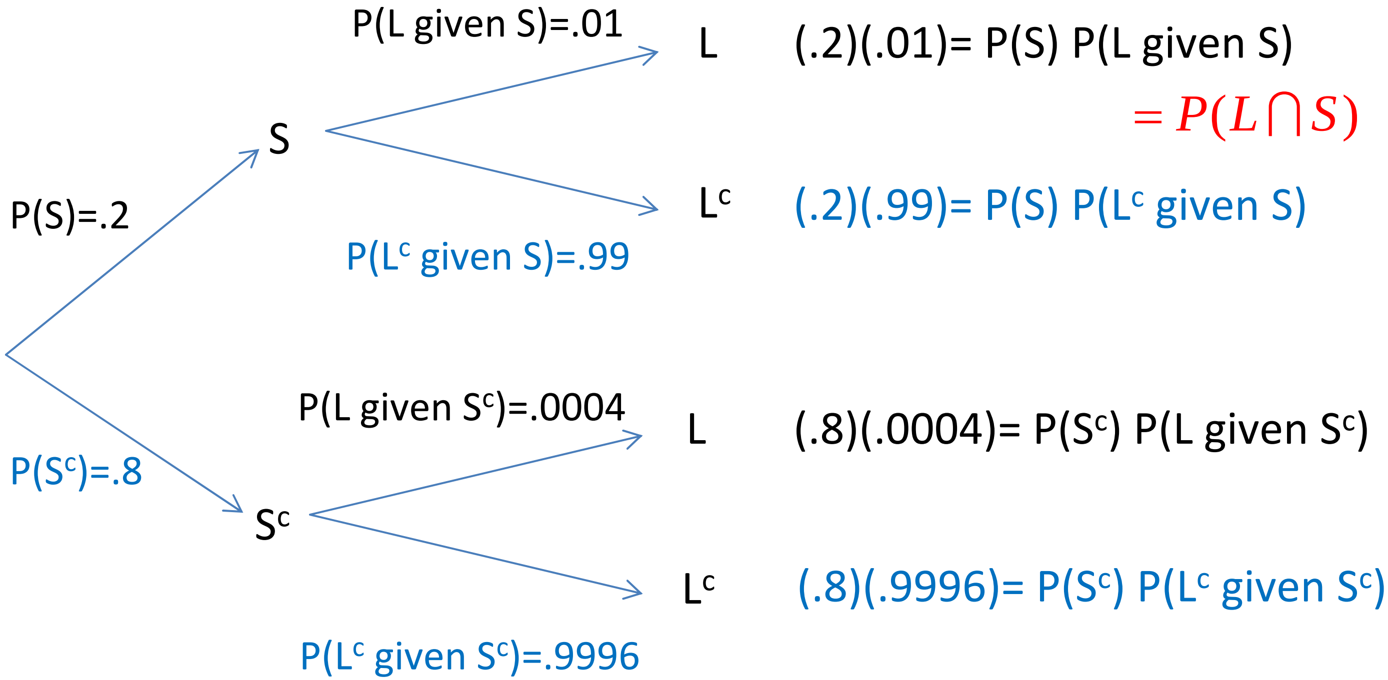


$(.2)(.01) = P(S)\ P(L \text{ given } S)$

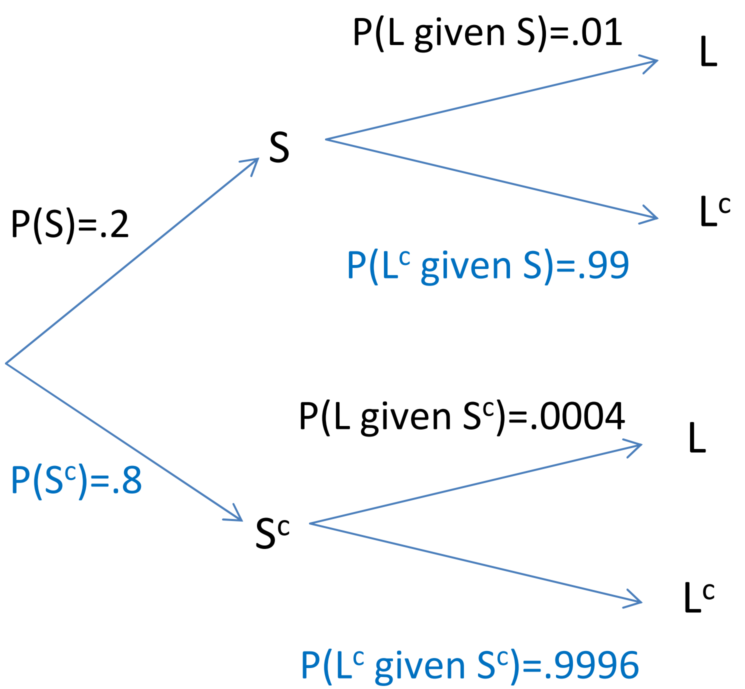
$(.2)(.99) = P(S)\ P(L^c \text{ given } S)$

$(.8)(.0004) = P(S^c)\ P(L \text{ given } S^c)$

$(.8)(.9996) = P(S^c)\ P(L^c \text{ given } S^c)$



P(L given S)=.01
L
(.2)(.01)= P(S) P(L given S)
$= P(L \cap S)$
S
P(S)=.2
L^c
$(.2)(.99)= P(S)\ P(L^c \text{ given } S)$
$P(L^c \text{ given } S)=.99$
$P(L \text{ given } S^c)=.0004$
L
$(.8)(.0004)= P(S^c)\ P(L \text{ given } S^c)$
$P(S^c)=.8$
S^c
L^c
$(.8)(.9996)= P(S^c)\ P(L^c \text{ given } S^c)$
$P(L^c \text{ given } S^c)=.9996$



(.2)(.01)= P(S) P(L given S)

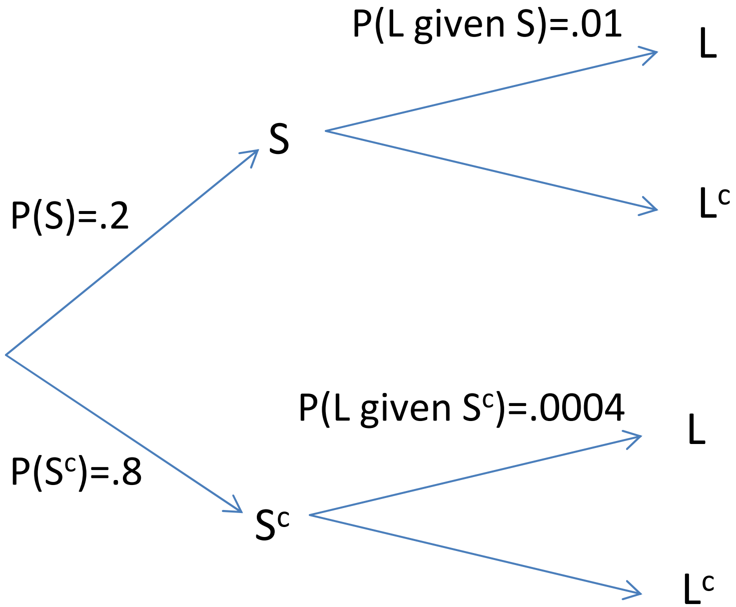
$$= P(L \cap S)$$

(.2)(.99)= P(S) P(L^c given S)

(.8)(.0004)= P(S^c) P(L given S^c)

$$= P(L \cap S^c)$$

(.8)(.9996)= P(S^c) P(L^c given S^c)



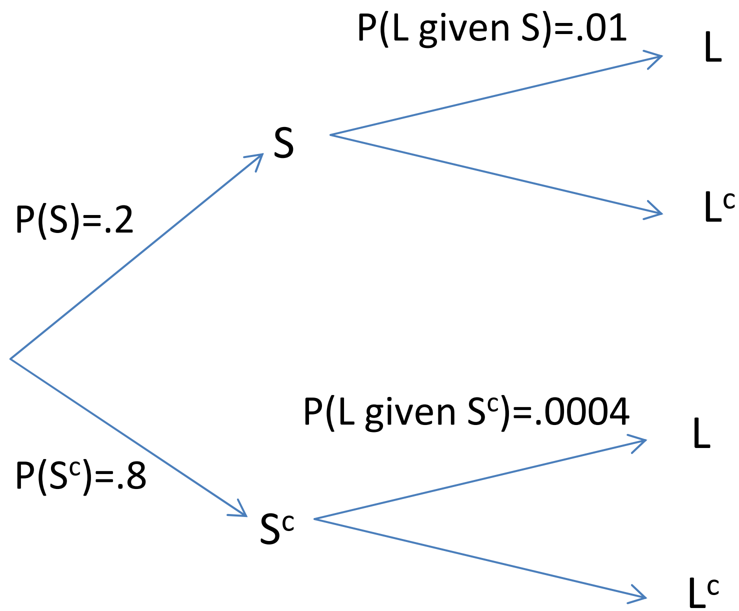
$(.2)(.01)= P(S)\ P(L \text{ given } S)$

$= P(L \cap S)$

$(.8)(.0004)= P(S^c)\ P(L \text{ given } S^c)$

$= P(L \cap S^c)$

What is P(S given L) ?



$(.2)(.01) = P(S)\ P(L \text{ given } S)$

$= P(L \cap S)$

$(.8)(.0004) = P(S^c)\ P(L \text{ given } S^c)$

$= P(L \cap S^c)$

What is P(S given L) ?

$$P(S\ given\ L) = \frac{P(S \cap L)}{P(L)} = \frac{P(L \cap S)}{P(L)}$$

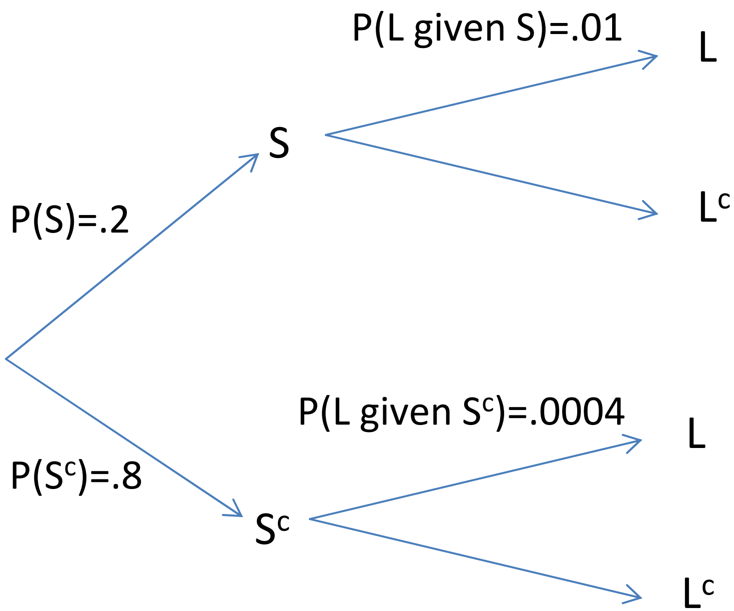
What is P(S given L) ?

$$P(S\ given\ L) = \frac{P(S \cap L)}{P(L)} = \frac{P(L \cap S)}{P(L)}$$

$$= \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)}$$

Since $L = (L \cap S) \cup (L \cap S^c)$

$$P(S \text{ given } L) = \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)}$$



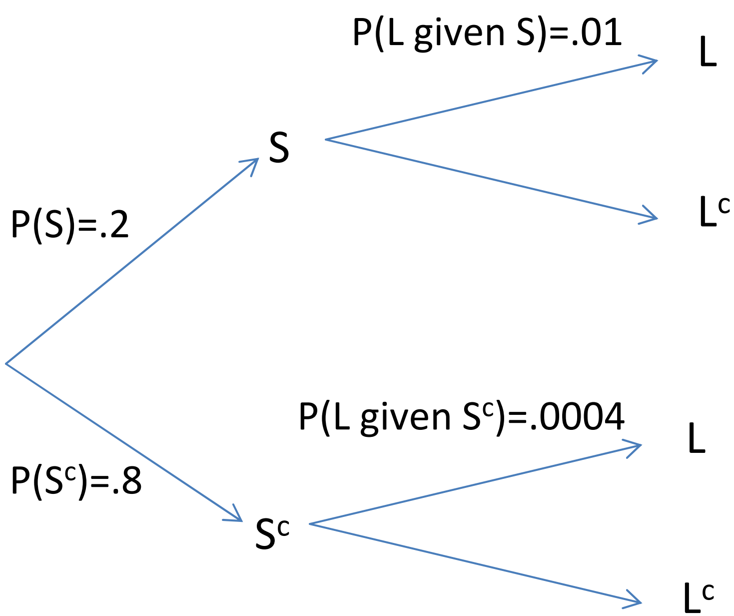
(.2)(.01)= P(S) P(L given S)

$= P(L \cap S)$

(.8)(.0004)= $P(S^c)$ P(L given S^c)

$= P(L \cap S^c)$

$$P(S\ given\ L) = \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)} = \frac{(.2)(.01)}{(.2)(.01) + (.8)(.0004)} \approx 0.86$$



$(.2)(.01) = P(S)\ P(L\ given\ S)$

$= P(L \cap S)$

$(.8)(.0004) = P(S^c)\ P(L\ given\ S^c)$

$= P(L \cap S^c)$

BAYES' THEOREM

version 1

$$P(S\ given\ L) = \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)}$$

BAYES' THEOREM

version 1

$$P(S\ given\ L) = \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)}$$

$$= \frac{P(S)P(L\ given\ S)}{P(S)P(L\ given\ S) + P(S^c)P(L\ given\ S^c)}$$

BAYES' THEOREM

version 2

If S_1, S_2, … , S_n are pairwise disjoint sets whose union is the entire sample space S, and if A is a subset of S, then,

$$P(S_1 \; given \; A) = \frac{P(S_1)P(A \, given \, S_1)}{P(S_1)P(A \, given \, S_1) + P(S_2)P(A \, given \, S_2) + \cdots + P(S_n)P(A \, given \, S_n)}$$

BAYES' THEOREM

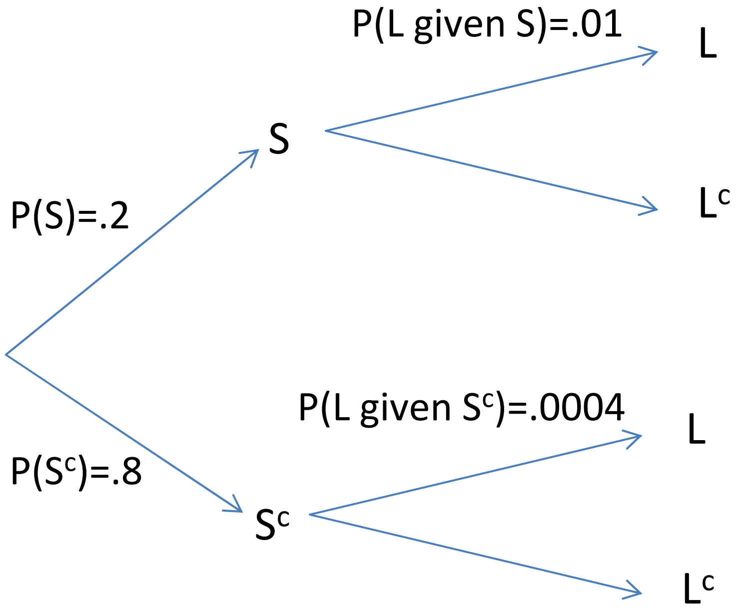
version 2

If S_1, S_2, ... , S_n are pairwise disjoint sets whose union is the entire sample space S, and if A is a subset of S, then,

$$P(S_1 \text{ given } A) = \frac{P(S_1)P(A \text{ given } S_1)}{P(S_1)P(A \text{ given } S_1) + P(S_2)P(A \text{ given } S_2) + \cdots + P(S_n)P(A \text{ given } S_n)}$$

$$= \frac{P(A \cap S_1)}{P(A \cap S_1) + P(A \cap S_2) + \cdots + P(A \cap S_n)}$$

$$P(S\ given\ L) = \frac{P(L \cap S)}{P(L \cap S) + P(L \cap S^c)} = \frac{(.2)(.01)}{(.2)(.01) + (.8)(.0004)} \approx 0.86$$



(.2)(.01)= P(S) P(L given S)

$= P(L \cap S)$

(.8)(.0004)= P(S^c) P(L given S^c)

$= P(L \cap S^c)$

Example 2, Page 402:

Registered voters in Marin County are 45% Democratic, 30% Republican, and 25% Independent. In the last election for county supervisor, 70% of the Democrats voted, as did 80% of the Republicans and 90% of the Independents. What is the probability that a randomly selected voter in this election was a Democrat?

Problem 9, Page 405:

In 1986 the Reagan administration issued an executive order allowing agency heads to subject all employees to urine tests for drugs. Suppose that the test is 95% accurate both in identifying drug users and in clearing nonusers. Suppose also that 1% of employees use drugs. What is the probability that a person who tests positive is not a drug user?