353 (disease) The incidence of a certain disease is one person in a thousand. There is a test for the disease that is 99% accurate. What is the probability that the person has the disease if they test

(a) positive (the test says they have the disease)?

(b) negative (the test says they don't have the disease)?

After trying the question, scroll down to the solution.

(a) positive (the test says they have the disease)?

§ Let d be a binary variable meaning that the person has the disease. We say that the incidence is one in a thousand as

$$\textbf{if } 0.001 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi}$$

The person tests positive, and the test is 99% accurate. This can be formalized as either

$$\textbf{if } d' \textbf{ then } 0.99 \textbf{ else } 0.01 \textbf{ fi}$$

or equivalently

$$\textbf{if } 0.99 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi}$$

And we can represent the two statements together by multiplying them together, and then normalizing (dividing by the sum). First we multiply.

$$\begin{aligned}
& \textbf{if } 0.001 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi} \times \textbf{if } 0.99 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi} \\
= \; & (0.001{\times}d' + 0.999{\times}(1{-}d')) \times (0.99{\times}d' + 0.01{\times}(1{-}d')) \\
= \; & 0.00999 + 0.96904{\times}d' - 0.97804{\times}d'^2
\end{aligned}$$

Now we normalize, which means dividing by

$$\begin{aligned}
& \Sigma d' \cdot 0.00999 + 0.96904{\times}d' - 0.97804{\times}d'^2 \\
= \; & 0.00999 + 0.96904{\times}0 - 0.97804{\times}0^2 + 0.00999 + 0.96904{\times}1 - 0.97804{\times}1^2 \\
= \; & 0.01098
\end{aligned}$$

So the two statements together are

$$(0.00999 + 0.96904{\times}d' - 0.97804{\times}d'^2) / 0.01098$$

And finally, the probability that the person has the disease is

$$\begin{aligned}
& (0.00999 + 0.96904{\times}d' - 0.97804{\times}d'^2) / 0.01098.\ d \\
= \; & \Sigma d'' \cdot (0.00999 + 0.96904{\times}d'' - 0.97804{\times}d''^2) / 0.01098 \times d'' \\
= \; & \quad (0.00999 + 0.96904{\times}0 - 0.97804{\times}0^2) / 0.01098 \times 0 \\
& + (0.00999 + 0.96904{\times}1 - 0.97804{\times}1^2) / 0.01098 \times 1 \\
= \; & 0.09 \text{ approximately}
\end{aligned}$$

There is approximately 9% chance the person has the disease. This calculation applies only if the person got tested randomly, for no reason. But people get tested for a disease when they have symptoms, and that changes things. For that more realistic situation, we would need to know the incidence of the disease in the subpopulation of people showing symptoms.

(b) negative (the test says they don't have the disease)?

§ Let d be a binary variable meaning that the person has the disease. We say that the incidence is one in a thousand as

$$\textbf{if } 0.001 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi}$$

The person tests negative, and the test is 99% accurate. This can be formalized as either

$$\textbf{if } 1{-}d' \textbf{ then } 0.99 \textbf{ else } 0.01 \textbf{ fi}$$

or equivalently

$$\textbf{if } 0.99 \textbf{ then } 1{-}d' \textbf{ else } d' \textbf{ fi}$$

And we can represent the two statements together by multiplying them together, and then normalizing (dividing by the sum). First we multiply.

$$\begin{aligned}
& \textbf{if } 0.001 \textbf{ then } d' \textbf{ else } 1{-}d' \textbf{ fi} \times \textbf{if } 0.99 \textbf{ then } 1{-}d' \textbf{ else } d' \textbf{ fi} \\
= \; & (0.001{\times}d' + 0.999{\times}(1{-}d')) \times (0.99{\times}(1{-}d') + 0.01{\times}d') \\
= \; & 0.98901 - 1.96704{\times}d' + 0.97804{\times}d'^2
\end{aligned}$$

Now we normalize, which means dividing by

$$\begin{aligned}
& \Sigma d' \cdot 0.98901 - 1.96704{\times}d' + 0.97804{\times}d'^2 \\
= \; & 0.98901 - 1.96704{\times}0 + 0.97804{\times}0^2 + 0.98901 - 1.96704{\times}1 + 0.97804{\times}1^2 \\
= \; & 0.98902
\end{aligned}$$

So the two statements together are

$(0.98901 - 1.96704 \times d' + 0.97804 \times d'^2) / 0.98902$

And finally, the probability that the person has the disease is

$(0.98901 - 1.96704 \times d' + 0.97804 \times d'^2) / 0.98902.\ d$

$= \quad \Sigma d'' \cdot (0.98901 - 1.96704 \times d'' + 0.97804 \times d''^2) / 0.98902 \times d''$

$= \quad (0.98901 - 1.96704 \times 0 + 0.97804 \times 0^2) / 0.98902 \times 0$

$\quad + (0.98901 - 1.96704 \times 1 + 0.97804 \times 1^2) / 0.98902 \times 1$

$= \quad 0.00001$ approximately

There is approximately 0.001% chance the person has the disease. This calculation applies only if the person got tested randomly, for no reason. But people get tested for a disease when they have symptoms, and that changes things. For that more realistic situation, we would need to know the incidence of the disease in the subpopulation of people showing symptoms.