PS 27: Problem 3.25

Case 1 1:100

Let $P(\text{D})$ represent the probability of having the disease given no other information, and $P(\text{ND})$ represent the probability of not having the disease. Since having the disease and not having the disease are mutually exclusive, we have

$$P(\text{D}) = \frac{1}{100} \tag{1}$$

$$P(\text{ND}) = 1 - P(\text{D}) = 1 - \frac{1}{100} = \frac{99}{100} \tag{2}$$

If the goal is for the test to have an accuracy of 50%, then we set the probability of having the disease given that you tested positive to $P(\text{D}|+) = 0.50$. The quantity of interest is the probability of testing positive given that you have the disease $P(+|\text{D})$. From Bayes' theorem,

$$P(\text{D}|+) = \frac{P(+|\text{D})P(\text{D})}{P(+|\text{D})P(\text{D}) + P(+|\text{ND})P(\text{ND})} \tag{3}$$

Plugging in (1) and (2),

$$\frac{50}{100} = \frac{P(+|\text{D})\frac{1}{100}}{P(+|\text{D})\frac{1}{100} + P(+|\text{ND})\frac{99}{100}} \tag{4}$$

The probability of testing positive given that you don't have the disease and the probability of testing positive given that you have the disease are also mutually exclusive. Therefore, we can write

$$\frac{50}{100} = \frac{P(+|\text{D})\frac{1}{100}}{P(+|\text{D})\frac{1}{100} + [1 - P(+|\text{D})]\frac{99}{100}} \tag{5}$$

Multiply both sides by 100 and solve for $P(+|\text{D})$.

$$\frac{50}{100} = \frac{P(+|\text{D})}{P(+|\text{D}) + 99[1 - P(+|\text{D})]}$$

$$\frac{50}{100}[P(+|\text{D}) + 99 - 99P(+|\text{D})] = P(+|\text{D})$$

$$P(+|\text{D})\left[\frac{50}{100} - \frac{50 \cdot 99}{100} - 1\right] = -\frac{50 \cdot 99}{100}$$

$$P(+|\text{D}) = \frac{-\frac{50}{100} \cdot 99}{\frac{50}{100} - \frac{50}{100} \cdot 99 - 1}$$

$$= \frac{-49.5}{-50}$$

$$\boxed{P(+|\text{D}) = 0.99} \tag{6}$$

Case 2 1:1000

With the same treatment as before we have the probabilities $P(\text{D}) = \frac{1}{1000}$ and $P(\text{ND}) = \frac{999}{1000}$. Following the same steps as in Case 1, we end up with the equation

$$P(+|\text{D}) = \frac{-\frac{50}{100} \cdot 999}{\frac{50}{100} - \frac{50}{100} \cdot 999 - 1}$$

$$= \frac{-499.5}{-500}$$

$$\boxed{P(+|\text{D}) = 0.999} \tag{7}$$

Case 3 1:10,000

Following the same arguments as before, the final equation is

$$P(+|\text{D}) = \frac{-\frac{50}{100} \cdot 9999}{\frac{50}{100} - \frac{50}{100} \cdot 9999 - 1}$$

$$= \frac{-4999.5}{-5000}$$

$$\boxed{P(+|\text{D}) = 0.9999} \tag{8}$$

Case 3 1:100,000

Following the same arguments as before, the final equation is

$$P(+|\text{D}) = \frac{-\frac{50}{100} \cdot 99,999}{\frac{50}{100} - \frac{50}{100} \cdot 99,999 - 1}$$

$$= \frac{-49,999.5}{-50,000}$$

$$\boxed{P(+|\text{D}) = 0.99999} \tag{9}$$

This shows that the more rare the occurrence is, the more unreasonably accurate the test has to be to achieve even just 50% true positive detection rate.