IMPORTANT: The answers are provided with the view to deepen your understanding of the topic and to stimulate more in depth discussions. They should not be shared with others or put on the web. The questions and answers may be revised in the next course delivery.

Answers to Tutorial 2

Qn 1

Let E be the event that an undergraduate student is male. Let F be the event that the student is in CENG.

$P(E) = 0.45$
$P(F) = 0.3$

a) $P(E|F) = 0.7$

b)
$$P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(E|F)P(F)}{P(E)} = \frac{(0.7)(0.3)}{0.45} \approx 0.4667$$

Qn 2

The Monte Hall problem is very difficult to accept even for eminent mathematicians. See https://en.wikipedia.org/wiki/Monty_Hall_problem.

One possible way:

Let E_i be the event that the contestant opens door i

Let F_j be the event that the host opens door j, where $j \neq i$

E_i and F_j are not independent. F_j depends on E_i. So the event F_j reveals some subtle information.

Qn 3

Readings

Qn 4

$$P(E|F) = \frac{(0.95)(10^{-6})}{(0.95)(10^{-6}) + (0.01)(1 - 10^{-6})} \approx 95 \times 10^{-6}$$

The chance is increased 95 times but it is still very small.

Observe that the false positive probability is $10^{-2} \gg 10^{-6}$, the probability of having the disease. So it suggests not to over-worry about very unlikely events.

<u>Qn 5</u>

Let E be the event that the patient has the cancer.
Let F be the event that the test indicates cancer, i.e., the test result is positive

Since the false negative rate is 1%,

$$P(F^c|E) = 0.01$$

Hence

$$P(F|E) = 0.99$$

Since the false positive rate is 10%

$$P(F|E^c) = 0.1$$

The doctor is 60% certain that the patient has the cancer. Thus

$$P(E) = 0.6$$

You may consider it as a subjective interpretation of probability. On the other hand, you may also interpret this objectively like this. The doctor has seen a large number of patients with the symptoms. The frequency for which patients having the symptoms have cancer is 0.6.

Note also that $P(F|E)$ describes the effectiveness of the test in identifying cancer when cancer is present. This probability is not related to the doctor's judgment.

$$P(E^c) = 1 - P(E) = 0.4$$

a)

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P(F|E)P(E)}{P(F)} = \frac{(0.99)(0.6)}{P(F)}$$

$P(F)$ is the probability that the test indicates cancer. Using Bayes' rule,

$$P(F) = P(F|E)P(E) + P(F|E^c)P(E^c) = (0.99)(0.6) + (0.1)(0.4) = 0.634$$

Hence

$$P(E|F) = 0.936908517 = 93.69\%$$

b)

$$P(E|F^c) = \frac{P(EF^c)}{P(F^c)} = \frac{P(F^c|E)P(E)}{1 - P(F)} = \frac{\big(1 - P(F|E)\big)P(E)}{1 - P(F)}$$

$$= \frac{(1 - 0.99)(0.6)}{(1 - 0.634)} = 0.016393442 \approx 1.64\%$$

c)

The test is useful in this situation. It is good at confirming the doctor's judgement (probability increases from 60% to more than 93%). On the other hand, if the test result is negative, it is quite reassuring (probability decreases from 40% to less than 2%).

Discuss: There may be situations when the test is not useful. Can you suggest one such situation?

Qn 6

Let E denote the event that the suspect is guilty and F the event that he possesses the characteristic of the criminal. We wish to calculate $P(E|F)$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$$P(EF) = P(F|E)P(E)$$

The inspector is 60% convinced that the suspect is the criminal. Hence $P(E) = 0.6$

As the new piece of evidence shows that the criminal has the characteristic,

$P(F|E) = 1$

Hence $P(EF) = (1)(0.6) = 0.6$

$P(F) = P(F|E)P(E) + P(F|E^c)P(E^c)$

As 20% of the population has this characteristic, $P(F|E^c) = 0.2$

$P(F) = (1)(0.6) + (0.2)(0.4) = 0.68$

Hence

$$P(E|F) = \frac{0.6}{0.68} = 0.882352941 \approx 0.882$$

Qn 7

Now $P(F|E) = 0.9$

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)} = \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)}$$

$$= 0.870967741 \approx 0.8710$$

Appendix: Terms related to "false positive" and "false negative"

#TP True Positive
#FP False Positive
#TN True Negative
#FN False Negative

$$Accuracy \quad = \frac{\#TP + \#TN}{\#TP + \#FP + \#TN + \#FN}$$

$$Sensitivity = \frac{\#TP}{\#TP + \#FN}$$

$$Specificity = \frac{\#TN}{\#TN + \#FP}$$

$$Precision \quad = \frac{\#TP}{\#TP + \#FP}$$

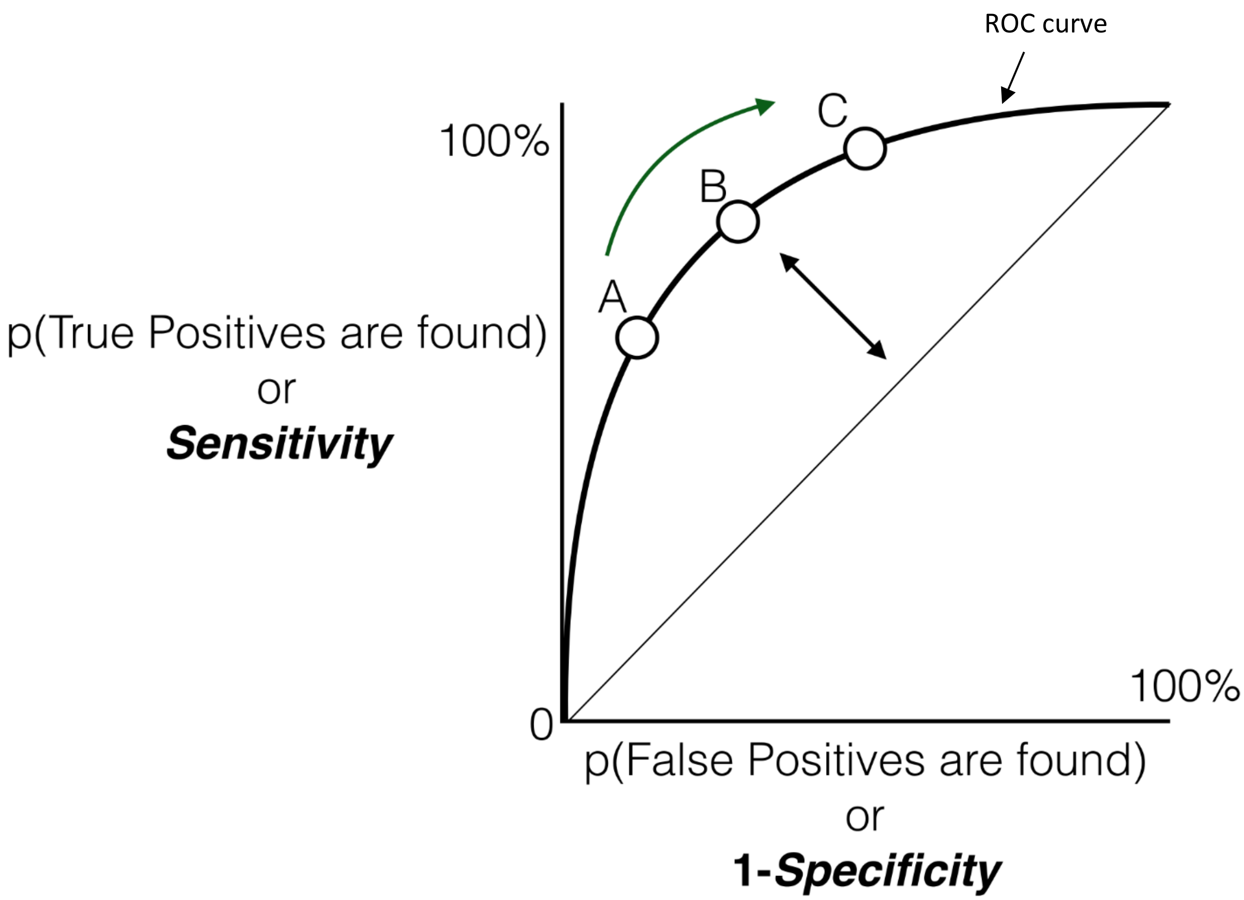
$$Recall \qquad = Sensitivity$$

Sensitivity (also known as True Positive Rate or Recall) is the probability that true positive is found.

Specificity is the probability that true negative is found.

1 – Precision is the false positive rate.

The performance of a classifier can be characterized by (Precision, Recall). Another popular way is the Receiver Operator Characteristic (ROC) curve, which plots the true positive rate against the false positive rate.



The diagonal line is a classifier using random guessing. Usually, when we improve the true positive rate, the false positive rate will also increase and vice versa (see the points from A to C).