BST 222 Homework 1 - Di Zhen

Problem 1

From the problem statement, we have

Sensitivity for test 1 = $P_1(+|D) = 84\%$, sensitivity for test 2 = $P_2(+|D) = 97\%$, specificity for test 1 = $P_1(-|\overline{D}) = 100\%$, specificity for test 2 = $P_1(-|\overline{D}) = 100\%$.

1. Suppose the prevalence of SARS-CoV-2 is 2%, i.e. $P(D) = 2\%$, then by Bayes Rule, the PPVs for test 1 $P_1(D|+)$ and test 2 $P_2(D|+)$ are

$$
\begin{aligned}
P_1(D|+) &= \frac{P_1(+,D)}{P_1(+)} \\
&= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + P_1(+|\overline{D})P(\overline{D})} \\
&= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + (1 - P_1(-|\overline{D}))(1 - P(D))} \\
&= \frac{84\% \times 2\%}{84\% \times 2\% + (1 - 100\%)(1 - 2\%)} \\
&= 1 \\
P_2(D|+) &= \frac{P_2(+,D)}{P_2(+)} \\
&= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + P_2(+|\overline{D})P(\overline{D})} \\
&= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + (1 - P_2(-|\overline{D}))(1 - P(D))} \\
&= \frac{97\% \times 2\%}{97\% \times 2\% + (1 - 100\%)(1 - 2\%)} \\
&= 1
\end{aligned}
$$

Suppose the prevalence of SARS-CoV-2 is 10%, i.e. $P(D) = 10\%$, then by Bayes Rule, the PPVs for test 1 $P_1(D|+)$ and test 2 $P_2(D|+)$ are

$$P_1(D|+) = \frac{P_1(+, D)}{P_1(+)}$$

$$= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + P_1(+|\overline{D})P(\overline{D})}$$

$$= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + (1 - P_1(-|\overline{D}))(1 - P(D))}$$

$$= \frac{84\% \times 10\%}{84\% \times 10\% + (1 - 100\%)(1 - 10\%)}$$

$$= 1$$

$$P_2(D|+) = \frac{P_2(+, D)}{P_2(+)}$$

$$= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + P_2(+|\overline{D})P(\overline{D})}$$

$$= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + (1 - P_2(-|\overline{D}))(1 - P(D))}$$

$$= \frac{97\% \times 10\%}{97\% \times 10\% + (1 - 100\%)(1 - 10\%)}$$

$$= 1$$

Suppose the prevalence of SARS-CoV-2 is 20%, i.e. $P(D) = 20\%$, then by Bayes Rule, the PPVs for test 1 $P_1(D|+)$ and test 2 $P_2(D|+)$ are

$$P_1(D|+) = \frac{P_1(+, D)}{P_1(+)}$$

$$= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + P_1(+|\overline{D})P(\overline{D})}$$

$$= \frac{P_1(+|D)P(D)}{P_1(+|D)P(D) + (1 - P_1(-|\overline{D}))(1 - P(D))}$$

$$= \frac{84\% \times 20\%}{84\% \times 20\% + (1 - 100\%)(1 - 20\%)}$$

$$= 1$$

$$P_2(D|+) = \frac{P_2(+, D)}{P_2(+)}$$

$$= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + P_2(+|\overline{D})P(\overline{D})}$$

$$= \frac{P_2(+|D)P(D)}{P_2(+|D)P(D) + (1 - P_2(-|\overline{D}))(1 - P(D))}$$

$$= \frac{97\% \times 20\%}{97\% \times 20\% + (1 - 100\%)(1 - 20\%)}$$

$$= 1$$

2. Suppose the prevalence of SARS-CoV-2 is 1%, i.e. $P(D) = 1\%$, then by Bayes Rule, the NPVs for test 1 $P_1(\overline{D}|-)$ and test 2 $P_2(\overline{D}|-)$ are

$$P_1(\overline{D}|-) = \frac{P_1(-,\overline{D})}{P_1(-)}$$

$$= \frac{P_1(-|\overline{D})P(\overline{D})}{P_1(-|D)P(D) + P_1(-|\overline{D})P(\overline{D})}$$

$$= \frac{P_1(-|\overline{D})(1 - P(D))}{(1 - P_1(+|D))P(D) + P_1(-|\overline{D})(1 - P(D))}$$

$$= \frac{100\% \times (1 - 1\%)}{(1 - 84\%) \times 1\% + 100\% \times (1 - 1\%)}$$

$$= 0.9984$$

$$P_2(\overline{D}|-) = \frac{P_2(-,\overline{D})}{P_2(-)}$$

$$= \frac{P_2(-|\overline{D})P(\overline{D})}{P_2(-|D)P(D) + P_2(-|\overline{D})P(\overline{D})}$$

$$= \frac{P_2(-|\overline{D})(1 - P(D))}{(1 - P_2(+|D))P(D) + P_2(-|\overline{D})(1 - P(D))}$$

$$= \frac{100\% \times (1 - 1\%)}{(1 - 97\%) \times 1\% + 100\% \times (1 - 1\%)}$$

$$= 0.9997$$

Suppose the prevalence of SARS-CoV-2 is 5%, i.e. $P(D) = 5\%$, then by Bayes Rule, the NPVs for test 1 $P_1(\overline{D}|-)$ and test 2 $P_2(\overline{D}|-)$ are

$$P_1(\overline{D}|-) = \frac{P_1(-,\overline{D})}{P_1(-)}$$

$$= \frac{P_1(-|\overline{D})P(\overline{D})}{P_1(-|D)P(D) + P_1(-|\overline{D})P(\overline{D})}$$

$$= \frac{P_1(-|\overline{D})(1 - P(D))}{(1 - P_1(+|D))P(D) + P_1(-|\overline{D})(1 - P(D))}$$

$$= \frac{100\% \times (1 - 5\%)}{(1 - 84\%) \times 5\% + 100\% \times (1 - 5\%)}$$

$$= 0.9916$$

$$P_2(\overline{D}|-) = \frac{P_2(-,\overline{D})}{P_2(-)}$$

$$= \frac{P_2(-|\overline{D})P(\overline{D})}{P_2(-|D)P(D) + P_2(-|\overline{D})P(\overline{D})}$$

$$= \frac{P_2(-|\overline{D})(1 - P(D))}{(1 - P_2(+|D))P(D) + P_2(-|\overline{D})(1 - P(D))}$$

$$= \frac{100\% \times (1 - 5\%)}{(1 - 97\%) \times 5\% + 100\% \times (1 - 5\%)}$$

$$= 0.9984$$

Suppose the prevalence of SARS-CoV-2 is 20%, i.e. $P(D) = 20\%$, then by Bayes Rule, the NPVs for test 1 $P_1(\overline{D}|-)$ and test 2 $P_2(\overline{D}|-)$ are

$$
\begin{aligned}
P_1(\overline{D}|-) &= \frac{P_1(-,\overline{D})}{P_1(-)} \\
&= \frac{P_1(-|\overline{D})P(\overline{D})}{P_1(-|D)P(D) + P_1(-|\overline{D})P(\overline{D})} \\
&= \frac{P_1(-|\overline{D})(1 - P(D))}{(1 - P_1(+|D))P(D) + P_1(-|\overline{D})(1 - P(D))} \\
&= \frac{100\% \times (1 - 20\%)}{(1 - 84\%) \times 20\% + 100\% \times (1 - 20\%)} \\
&= 0.9615 \\
P_2(\overline{D}|-) &= \frac{P_2(-,\overline{D})}{P_2(-)} \\
&= \frac{P_2(-|\overline{D})P(\overline{D})}{P_2(-|D)P(D) + P_2(-|\overline{D})P(\overline{D})} \\
&= \frac{P_2(-|\overline{D})(1 - P(D))}{(1 - P_2(+|D))P(D) + P_2(-|\overline{D})(1 - P(D))} \\
&= \frac{100\% \times (1 - 20\%)}{(1 - 97\%) \times 20\% + 100\% \times (1 - 20\%)} \\
&= 0.9926
\end{aligned}
$$

3. There is 84% probability that diseased people will be tested positive and 16% probability that the diseased people will be tested negative. If one is not diseased, he/she is impossible to be tested positive. If one is tested positive, it is impossible that he/she is healthy. If one is tested negative, the probability of not being diseased increases with the increased proportion of patients in the population. In other words, if the patient is tested negative, there is a small probability that the he/she is still diseased, and this probability is getting smaller if the proportion of patients in the population is larger.

Problem 2

Supposed that vaccinated $= V$, exposed $= E$, from the problem statement we know that

$P(\overline{D}|V, E) = 70\%, \quad P(D|\overline{V}, E) = 50\%.$

Since those three people are independent, the probability that at least one will get the disease if all were exposed is equal to

$$
1 - P(\overline{D}|V, E) \cdot P(\overline{D}|\overline{V}, E) \cdot P(\overline{D}|V, E) = 1 - 0.7 \times (1 - 0.5) \times 0.7 = 0.755
$$

Thus, the probability that at least one will get the disease if all were exposed is 0.755.

Let X denotes the number of diseased people, then X can be $0, 1, 2, 3$.

$$P(X = 0) = P(\overline{D}|V)^2 P(\overline{D}|\overline{V}) = 0.7^2 \times 0.5 = 0.245$$
$$P(X = 1) = 2 \times P(D|V)P(\overline{D}|V)P(\overline{D}|\overline{V}) + P(D|\overline{V})P(\overline{D}|V)^2 = 0.455$$
$$P(X = 2) = P(D|V)^2 P(\overline{D}|\overline{V}) + P(\overline{D}|V)P(D|V)P(D|\overline{V}) + P(D|V)P(\overline{D}|V)P(D|\overline{V}) = 0.255$$
$$P(X = 3) = P(D|V)^2 P(D|\overline{V}) = 0.3^2 \times 0.5 = 0.045$$

Therefore, the PMF is

$$P(X = x) = \begin{cases} 0.245, & x = 0 \\ 0.455, & x = 1 \\ 0.255, & x = 2 \\ 0.045, & x = 3 \end{cases}$$

The CDF is

$$F(X \leq x) = \begin{cases} 0.245, & x \in (0, 1] \\ 0.7, & x \in (1, 2] \\ 0.955, & x \in (2, 3] \\ 1, & x \in (3, 4] \end{cases}$$

Problem 3

We can estimate $\pi_{11}, \pi_{01}, \pi_{10}$ as follows

$$\pi_{11} = \theta_{11}p_1^2 + \theta_{10}p_1p_2 + \theta_{01}p_1p_2 + (1 - \theta_{11} - \theta_{01} - \theta_{10})p_2^2$$
$$\pi_{10} = \theta_{11}p_1p_2 + \theta_{10}p_1p_1 + \theta_{01}p_2^2 + (1 - \theta_{11} - \theta_{01} - \theta_{10})p_2p_1$$
$$\pi_{01} = \theta_{11}p_2p_1 + \theta_{10}p_2^2 + \theta_{01}p_1^2 + (1 - \theta_{11} - \theta_{01} - \theta_{10})p_1p_2$$

Problem 4

$$\int_0^1 f(x)dx = \int_0^1 kx^2(1 - x)dx$$

1.

$$= k\int_0^1 (x^2 - x^3)dx$$
$$= k\left(\frac{1}{3}x^3 - \frac{1}{4}x^4\right)\Big|_0^1$$
$$= k\left(\frac{1}{3} - \frac{1}{4} - 0\right)$$
$$= \frac{k}{12} = 1$$

So $k = 12$, and the function is $f(x) = 12x^2(1 - x)$.

$$E[X] = \int_0^1 x\,12x^2(1-x)dx$$

$$= 12\int_0^1 (x^3 - x^4)dx$$

2.
$$= 12\left(\frac{1}{4}x^4 - \frac{1}{5}x^5\right)\Big|_0^1$$

$$= 12\left(\frac{1}{4} - \frac{1}{5}\right)$$

$$= \frac{3}{5}$$

Problem 5

A function $g(x)$ is a PDF iff $g(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} g(x)dx = 1$.

Since $f(x)$ is a PDF, $f(x) \geq 0$. Since $F(x)$ is a CDF, for any $x_0 \leq x$, $0 \leq F(x_0) < 1$. So $g(x) \geq 0$.

In addition,

$$\int_{-\infty}^{\infty} g(x)dx = \int_{x_0}^{\infty} \frac{f(x)}{[1 - F(x_0)]}dx = \frac{1 - F(x_0)}{1 - F(x_0)} = 1$$

Therefore, $g(x)$ is a PDF.

Problem 6

1. Applying Theorem 2.1.5, first when $g(x) = x^2$, calculate

$$g^{-1}(y) = \sqrt{y}, \quad \frac{dg^{-1}(y)}{dy} = \frac{1}{2}\frac{1}{\sqrt{y}}$$

So

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1$$

2. Applying Theorem 2.1.5, first when $g(x) = -\log x$, calculate

$$g^{-1}(y) = e^{-y}, \quad \frac{dg^{-1}(y)}{dy} = -e^{-y}$$

So

$$f_Y(y) = f_X(e^{-y})e^{-y} = \frac{(n+m+1)!}{n!m!}(e^{-y(n+1)})(1 - e^{-y})^m, \quad 0 < y < \infty$$

3. Applying Theorem 2.1.5, first when $g(x) = e^x$, calculate

$$g^{-1}(y) = \log y, \quad \frac{dg^{-1}(y)}{dy} = \frac{1}{y}$$

So

$$f_Y(y) = f_X(\log y)\frac{1}{y} = \frac{1}{\sigma^2}\frac{1}{y}\log y \, e^{-(\log y/\sigma)^2/2}, \quad 0 < y < \infty$$