CS 215 – Assignment 2

Challa Siva Ramya-24B0941
Gunda Joshmitha-24B1098

1 Problem 1

Given n subjects,

1.1 a

The number of pools is $\frac{n}{s}$, and the number of subjects in each pool is s.

i. Number of tests in round 1 is $\frac{n}{s}$, since each pool is tested once.

   In round 2: The probability that a pool tests positive in the first round is

$$1 - (1-p)^s,$$

   because at least one of the s subjects should be diseased.

   Thus, for a pool with s subjects, the expected number of tests is

$$s \cdot (1 - (1-p)^s),$$

   and for $\frac{n}{s}$ pools, the expected number of tests is

$$\tfrac{n}{s} \cdot s \cdot (1 - (1-p)^s) = n(1 - (1-p)^s).$$

   Therefore, the total number of tests is

$$T(s) = \frac{n}{s} + n\big(1 - (1-p)^s\big).$$

ii. If p is very small, then $(1-p)^s \approx 1 - ps$, so

$$T(s) \approx \frac{n}{s} + nps.$$

   Differentiating:

$$T'(s) = -\frac{n}{s^2} + np.$$

   Setting $T'(s) = 0$ gives

$$\frac{n}{s^2} = np \quad \Longrightarrow \quad s^2 = \frac{1}{p}, \quad s = \frac{1}{\sqrt{p}}.$$

   Hence, the expected value of $T(s)$ at optimum is

$$T(s) = 2n\sqrt{p}.$$

iii. For $T(s) < n$:

$$\frac{n}{s} + n(1 - (1-p)^s) < n \quad \Longrightarrow \quad \frac{1}{s} < (1-p)^s.$$

Equivalently,

$$p < 1 - \sqrt[s]{\tfrac{1}{s}}.$$

Let $f(s) = 1 - \sqrt[s]{\frac{1}{s}}$. Then p must be less than the maximum of $f(s)$.

Differentiating $f(s)$ shows the maximum occurs near $s = e$. Since s must be a natural number, checking $s = 2$ and $s = 3$ gives the maximum at $s = 3$, with value ≈ 0.306.

Therefore, the greatest value of p for which $T(s) < n$ is about 0.306.

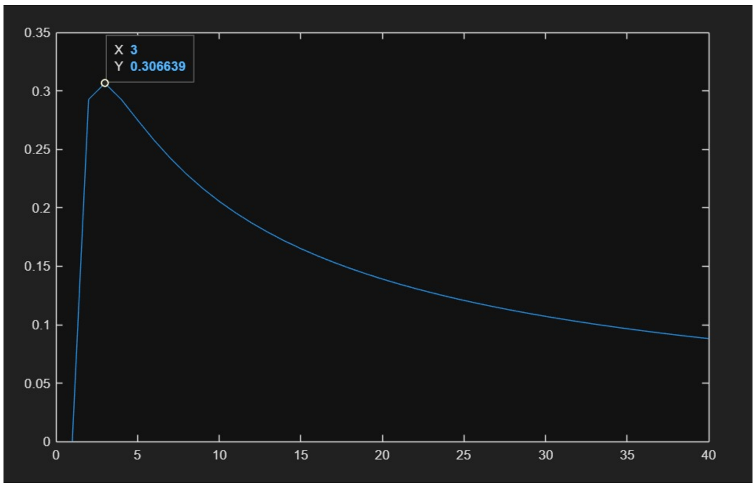

Figure 1: Plot of p vs s

```
For s = 2,f=0.2929
For s = 3,f=0.3066
the larger f is 0.3066 , which occurs at s=3.
```

Figure 2: Maximum value of p at $s = 3$.

1.2 b

i. For a particular pool, the probability that a subject joins is π and that they have the disease is p, so the probability a person joining is diseased is $p\pi$.

For the pool to test negative, all subjects in it must be healthy. For the other $n-1$ members this probability is $(1-p\pi)^{n-1}$. For the given healthy subject to join, the probability is π.

Hence the required probability is

$$\pi \cdot (1 - p\pi)^{n-1}.$$

ii. Let
$$f(\pi) = \pi(1 - p\pi)^{n-1}.$$
Differentiating,
$$f'(\pi) = (1 - p\pi)^{n-1} - p\pi(n-1)(1 - p\pi)^{n-2}.$$
Simplifying,
$$f'(\pi) = (1 - np\pi)(1 - p\pi)^{n-2}.$$
Thus, $f'(\pi) = 0$ at $\pi = \frac{1}{np}$ (since $\pi = \frac{1}{p}$ is infeasible).

Checking the second derivative shows it is a maximum. Therefore, the probability is maximized when
$$\pi = \frac{1}{np}.$$

iii. From part (i), the probability that a pool tests negative given a healthy subject participates is
$$\pi(1 - p\pi)^{n-1}.$$
So, the probability it tests positive is
$$1 - \pi(1 - p\pi)^{n-1}.$$
The probability that all pools a healthy subject participates in test negative is
$$\left(\pi(1 - p\pi)^{n-1}\right)^{T_1}.$$
Substituting $\pi = \frac{1}{np}$,
$$\left(\frac{(n-1)^{n-1}}{p\,n^n}\right)^{T_1}.$$

iv. Total expected number of tests = round 1 + round 2.

Round 1: T_1 tests.

Round 2:

- Diseased subjects: probability they appear in a pool $= 1 - (1-\pi)^{T_1}$. Expected tests $= np\,(1 - (1-\pi)^{T_1})$.
- Healthy subjects: probability they appear in a positive pool $= 1 - (1 - \pi(1 - p\pi)^{n-1})^{T_1}$. Expected tests $= n(1-p)\left[1 - (1 - \pi(1 - p\pi)^{n-1})^{T_1}\right]$.

Therefore,
$$\mathbb{E}[\text{Total tests}] = T_1 + np\big(1 - (1-\pi)^{T_1}\big) + n(1-p)\Big(1 - (1 - \pi(1-p\pi)^{n-1})^{T_1}\Big).$$

v. To minimize, take $\pi = \frac{1}{np}$ (from (ii)), and then choose T_1 that minimizes the above expression.

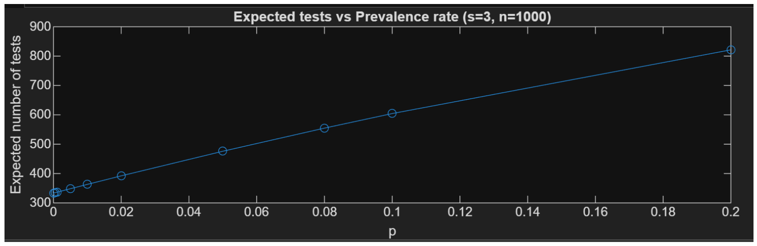

Figure 3: Method 1

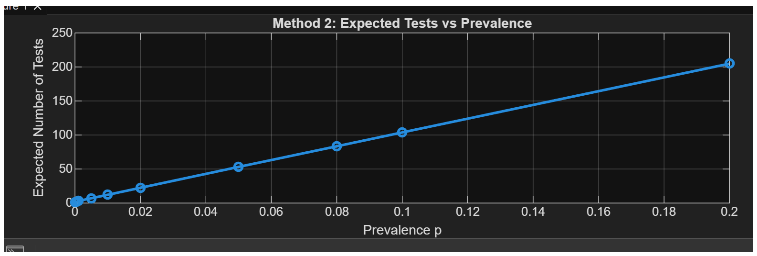

Figure 4: Method 2

1.3 c

Comments

At low prevalence p the number of expected tests is much smaller and as p increases the curve rises almost linearly so method 1 works well for rare diseases but not for diseasese that are not rare.
in method 2 the number of tests is much smaller than method 1 for all values of p the growth is linear but with a smaller slope , hinting its more efficient.

2 Problem 2

We start with

$$F_Z(z) = P(Z \le z) = P(XY \le z).$$

Case 1: $a > 0$

$$P(XY \le z \mid X = a) = P\big(Y \le \tfrac{z}{a}\big) = F_Y\big(\tfrac{z}{a}\big).$$

Contribution:

$$f_X(a)\,da \cdot F_Y\big(\tfrac{z}{a}\big).$$

Case 2: $a < 0$

$$P(XY \le z \mid X = a) = P\big(Y \ge \tfrac{z}{a}\big) = 1 - F_Y\big(\tfrac{z}{a}\big).$$

Contribution:

$$f_X(a)\,da \cdot \big(1 - F_Y(\tfrac{z}{a})\big).$$

Case 3: $a = 0$

If $X = 0$, then $Z = 0$ with probability 1:

$$F_Z(z) = \begin{cases} P(X = 0), & z \geq 0, \\ 0, & z < 0. \end{cases}$$

Let this contribution be C.

Thus

$$F_Z(z) = \int_{-\infty}^{0} \big(1 - F_Y(\tfrac{z}{a})\big) f_X(a)\, da + \int_{0}^{\infty} F_Y(\tfrac{z}{a}) f_X(a)\, da + C.$$

Differentiating,

$$f_Z(z) = \int_{-\infty}^{0} \frac{-1}{a} f_Y\left(\tfrac{z}{a}\right) f_X(a)\, da + \int_{0}^{\infty} \frac{1}{a} f_Y\left(\tfrac{z}{a}\right) f_X(a)\, da.$$

Final Expression:

$$f_Z(z) = \int_{-\infty}^{0} \frac{-1}{a} f_Y\left(\tfrac{z}{a}\right) f_X(a)\, da + \int_{0}^{\infty} \frac{1}{a} f_Y\left(\tfrac{z}{a}\right) f_X(a)\, da.$$

3 Problem 3

We are asked whether the estimator of $\mathbb{E}[X]$ should be

$$\hat{x}_1 := \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{or} \quad \hat{x}_2 := \frac{1}{n} \sum_{i=1}^{n} f_X(x_i)\, x_i.$$

Case 1: $\hat{x}_1$

$$\mathbb{E}[\hat{x}_1] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} x_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[x_i] = \mathbb{E}[X].$$

Thus, $\hat{x}_1$ is an unbiased estimator.

Case 2: $\hat{x}_2$

$$\mathbb{E}[\hat{x}_2] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[f_X(x_i)\, x_i] = \mathbb{E}[f_X(X)\, X].$$

This is not equal to $\mathbb{E}[X]$. Hence $\hat{x}_2$ is biased.

Conclusion

The correct estimator is

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

4 Problem 4

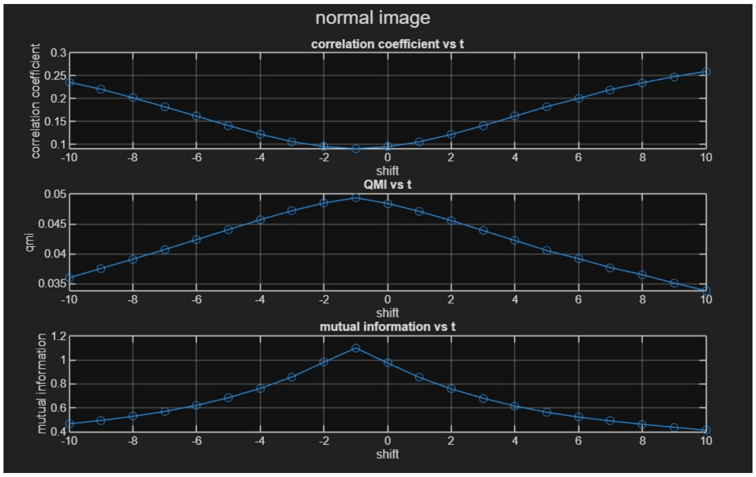

Figure 5: Normal image: QMI and MI peak near zero shift, showing better alignment than correlation.

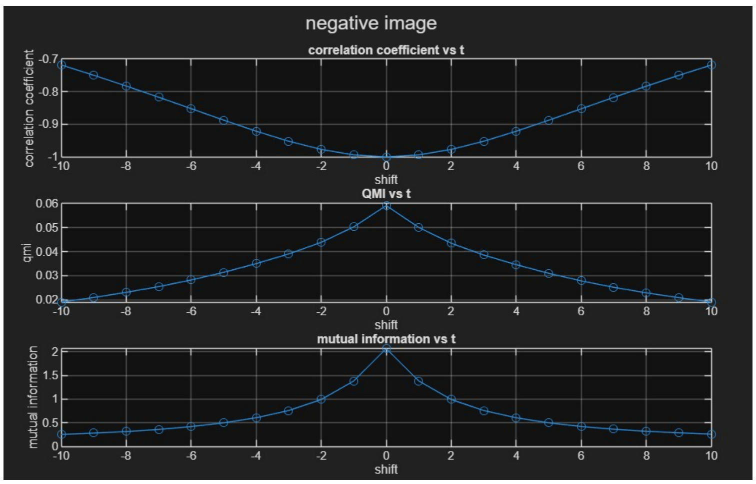

Figure 6: Negative image: correlation is low at zero shift, but QMI and MI still peak at zero, showing similarity.

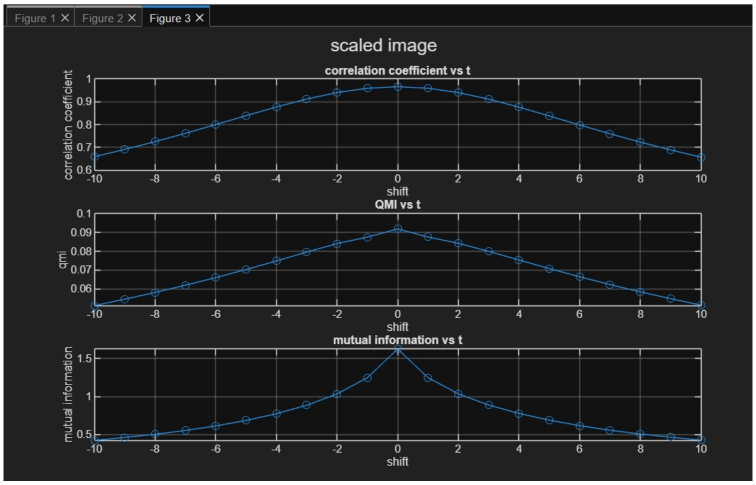

Figure 7: Scaled image: all measures peak near zero shift, showing alignment, with smooth symmetric curves.

5 Problem 5

(i)

We want to show

$$\mathbb{P}(X \geq x) \leq e^{-tx}\,\varphi_X(t), \quad t > 0.$$

By Markov's inequality,

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} = e^{-tx}\,\varphi_X(t).$$

(ii)

Similarly, for $t < 0$,

$$\mathbb{P}(X \leq x) \leq e^{-tx}\,\varphi_X(t).$$

(iii)

For $X = X_1 + \cdots + X_n$, with $\mathbb{E}[X] = \mu$,

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp\big(\mu(e^t - 1) - t(1+\delta)\mu\big).$$

(iv)

Optimizing at $t = \ln(1+\delta)$ gives

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp\Big(\mu\delta - \mu(1+\delta)\ln(1+\delta)\Big).$$

6 Problem 6

Let T be the trial number of the first head. Then

$$\mathbb{P}(T = t) = (1-p)^{t-1}p, \quad t = 1, 2, \ldots$$

If there are n tosses,

$$\mathbb{E}[T] = \sum_{t=1}^{n} t(1-p)^{t-1}p + (n+1)(1-p)^n.$$

Using the sum identity,

$$\sum_{t=1}^{n} t(1-p)^{t-1} = \frac{1-(n+1)(1-p)^n + n(1-p)^{n+1}}{p^2}.$$

Therefore,

$$\mathbb{E}[T] = \frac{1}{p}\Big(1 + (1-p)^n(p-n)\Big).$$