CSE546 - Homework # 0 - Solutions

Cheng-Yen Yang

April 7, 2021

Problem A.1

From the problem statements, we know:

$$P(positive|disease) = 0.99, P(positive|no\ disease) = 0.01, \tag{1}$$

and

$$P(disease) = 0.0001, P(no\ disease) = 0.9999. \tag{2}$$

By using Bayes Rule, we can obtained:

$$\begin{align}
P(disease|positive) &= \frac{P(positive|disease)P(disease)}{P(positive)} \tag{3}\\
&= \frac{P(positive|disease)P(disease)}{P(positive|disease)P(disease) + P(positive|no\ disease)P(no\ disease)} \tag{4}\\
&= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.0001 \times 0.9999} \tag{5}\\
&= \frac{1}{102} \tag{6}
\end{align}$$

that the chances of having the disease if tested positive is approximately $1/102$.

Problem A.2

a.

First we derive the *covariance* defined by the problem using the linearity of expectation:

$$\begin{align}
Cov(X, Y) &= \big[(X - E[X])(Y - E[Y])\big] \tag{7}\\
&= E[X \cdot Y] - E[X \cdot E[Y]] - E[Y \cdot E[X]] + E[X] \cdot E[Y] \tag{8}\\
&= E[X \cdot Y] - E[X] \cdot E[Y]. \tag{9}
\end{align}$$

Then by the law of total expectation:

$$\begin{align}
E[X \cdot Y] &= \sum_{x \in X} E[X \cdot Y | X = x] \cdot P(X = x) \tag{10}\\
&= \sum_{x \in X} x \cdot E[Y | X = x] \cdot P(X = x) \tag{11}\\
&= \sum_{x \in X} x^2 \cdot P(X = x) \tag{12}\\
&= E[X^2], \tag{13}
\end{align}$$

and

$$
\begin{aligned}
E[X] \cdot E[Y] &= E[X] \cdot \sum_{x \in X} E[Y|X=x] \cdot P(X=x) && (14)\\
&= E[X] \cdot \sum_{x \in X} x \cdot P(X=x) && (15)\\
&= E[X] \cdot E[X] && (16)\\
&= \big(E[X]\big)^2 && (17)
\end{aligned}
$$

By inserting the formations from (13) and (17) into (9), we have:

$$
Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y] = E[X^2] - \big(E[X]\big)^2 = E\big[(X - E[X])^2\big]. \tag{18}
$$

b.

If X and Y are independent, we re-derive the term $E[X \cdot Y]$:

$$
\begin{aligned}
E[X \cdot Y] &= \sum_{x \in X, y \in Y} E[X \cdot Y|X=x, Y=y] \cdot P(X=x, Y=y) && (19)\\
&= \sum_{x \in X, y \in Y} x \cdot y \cdot P(X=x) \cdot P(Y=y) && (20)\\
&= \sum_{x \in X} x \cdot P(X=x) \sum_{y \in Y} y \cdot P(Y=y) && (21)\\
&= E[X] \cdot E[Y], . && (22)
\end{aligned}
$$

By inserting the formations from (22) into (9), we have:

$$
Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y] = 0. \tag{23}
$$

Problem A.3

a.

From the problem statements we have:

$$
\begin{aligned}
H(Z) &= P(Z \leq z) && (24)\\
&= P(X + Y \leq z) && (25)\\
&= \int\int_{x+y \leq z} f(x) \cdot g(y) \cdot dx \cdot dy && (26)\\
&= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{z-x} g(y) \cdot dy \Big) f(x) \cdot dx && (27)\\
&= \int_{-\infty}^{\infty} G(z-x) \cdot f(x) \cdot dx, && (28)
\end{aligned}
$$

then by $h(z) = \frac{d}{dx} H(Z)$, we derive:

$$h(z) = \frac{d}{dx} H(Z) \tag{29}$$

$$= \frac{d}{dx} \int_{-\infty}^{\infty} G(z-x) \cdot f(x) \cdot dx \tag{30}$$

$$= \int_{-\infty}^{\infty} f(x) \cdot g(z-x) \cdot dx \tag{31}$$

b.

From Problem A.3.a, we have:

$$h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x) \cdot dx, \tag{32}$$

then if X and Y are both independent and uniformly distributed on $[0, 1]$, we have:

$$h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x) \cdot dx \tag{33}$$

$$= \int_{0}^{1} g(z-x) \cdot dx \tag{34}$$

$$= \int_{0}^{1} g(x)_{x \in (-\infty, z)} dx + g(x)_{x \in (z-1, \infty)} dx \tag{35}$$

$$= z_{z \in [0,1]} + (2-z)_{z \in [1,2]}, \tag{36}$$

therefore.

$$h(z) = \begin{cases} z, & z \in [0, 1) \\ 1, & z = 1 \\ 2 - z, & z \in (1, 2] \\ 0, & \text{otherwise} \end{cases} \tag{37}$$

Problem A.4

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$ are both Gaussian, to have:

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2) = \mathcal{N}(0, 1), \tag{38}$$

we need to assign a and b as:

$$a = \frac{1}{\sigma}, \text{ and } b = -\frac{\mu}{\sigma}. \tag{39}$$

Problem A.5

First we compute the mean of $\sqrt{n}(\widehat{\mu}_n - \mu)$:

$$E\Big[\sqrt{n}(\widehat{\mu}_n - \mu)\Big] = \sqrt{n} E\Big[(\widehat{\mu}_n - \mu)\Big] \tag{40}$$

$$= \sqrt{n} E\Big[\frac{1}{n} \sum_{i=1}^{n} X_i - \mu)\Big] \tag{41}$$

$$= \sqrt{n} E\Big[\frac{1}{n}(n\mu) - \mu)\Big] \tag{42}$$

$$= 0. \tag{43}$$

Then we compute the variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$:

$$\begin{aligned}
Var\Big[\sqrt{n}(\widehat{\mu}_n - \mu)\Big] &= nVar\Big[(\widehat{\mu}_n - \mu)\Big] && (44)\\
&= nVar\Big[(\widehat{\mu}_n)\Big] && (45)\\
&= nVar\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] && (46)\\
&= n\frac{\sigma^2}{n} && (47)\\
&= \sigma^2. && (48)
\end{aligned}$$

Problem A.6

a.

For any x, we have:

$$\begin{aligned}
E[\widehat{F}_n(x)] &= E\Big[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}\Big] && (49)\\
&= \frac{1}{n}\sum_{i=1}^{n} E[\mathbf{1}\{X_i \le x\}] && (50)\\
&= \frac{1}{n}\Bigg(E[X_1 \le x] + E[X_2 \le x] + \cdots + E[X_n \le x]\Bigg), && (51)
\end{aligned}$$

then from the problem statement we knew that $X_1, \cdots, X_n$ are independent and identically distributed random variables with the CDF $F(x)$. Therefore, we have:

$$E[\widehat{F}_n(x)] = \frac{1}{n} \cdot \big(n \cdot F(x)\big) = F(x) \quad (52)$$

b.

For any x, we have:

$$\begin{aligned}
Var[\widehat{F}_n(x)] &= Var\Big[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}\Big] && (53)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\Big[\mathbf{1}\{X_i \le x\}\Big] && (54)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} \Bigg(E\Big[(\mathbf{1}\{X_i \le x\})^2\Big] - \Big[\mathbf{1}\{X_i \le x\}\Big]\Big)^2\Bigg), && (55)
\end{aligned}$$

then from (52) we have:

$$Var[\widehat{F}_n(x)] = \frac{1}{n^2} \cdot \big(n \cdot (F(x) - F(x)^2)\big) = \frac{F(x)(1 - F(x))}{n} \quad (56)$$

c.

First, from Problem A.6.a and A.6.b we have:

$$E\Big[\big(\widehat{F}_n(x) - F(x)\big)^2\Big] = Var\Big[\widehat{F}_n(x) - F(x)\Big] = \frac{F(x)(1-F(x))}{n}, \tag{57}$$

as $E\Big[\widehat{F}_n(x) - F(x)\Big] = 0$ and since $F(x)$ is the CDF of some independent and identically distributed random variables, we know that:

$$E\Big[\big(\widehat{F}_n(x) - F(x)\big)^2\Big] = \frac{F(x)(1-F(x))}{n} \leq \frac{1}{4n}, \tag{58}$$

given that $F(x) \in [0,1]\ \forall\ x$.

Problem A.7

a.

We first compute the R.R.E.F. of matrix A and B:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{59}$$

and

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \longrightarrow = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{60}$$

Therefore,

$$rank(A) = 2, \text{ and } rank(B) = 2. \tag{61}$$

b.

We know that a leading one is the first nonzero entry in a row. The columns containing leading ones are the pivot columns. To obtain a basis for the column space, we just use the pivot columns from the original matrix and can find that:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\} \tag{62}$$

is a minimal size basis for A and B's column space.

Problem A.8

a.

$$Ac = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0+2+4 \\ 2+4+2 \\ 3+3+1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix} \tag{63}$$

b.

To solve the linear system $Ax = b$, We have the augmented matrix $[A|b]$:

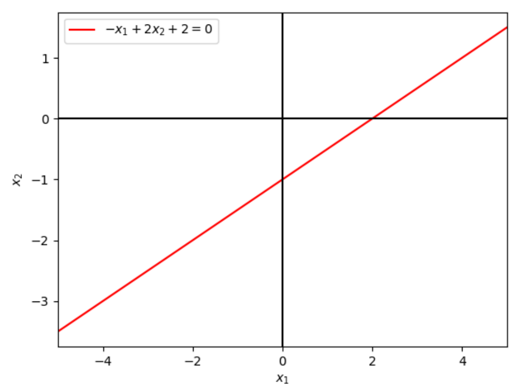
$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 2 & 1 & -1 \\ 3 & 3 & 1 & -4 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 3 & 3 & 1 & -4 \end{array}\right] \tag{64}$$

$$\longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & -3 & -2 & -1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right] \tag{65}$$

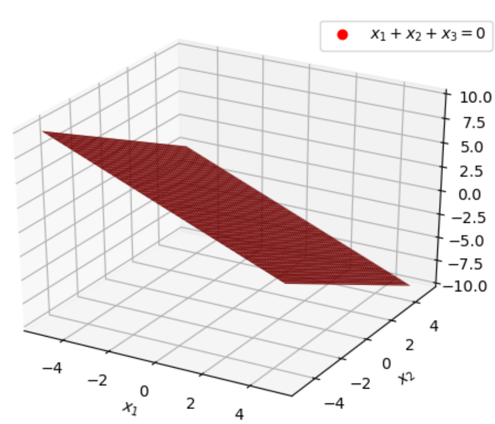
we obtained: $x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$.

Problem A.9

a.



b.



c.

From the hint, we choose $\widehat{x}_0$ to be the minimizer, the nearest point to x_0 on the hyper-plane $\omega^T x + b = 0$ which gives us:

$$\|x_0 - \widehat{x}\| = \left|\frac{w^T(x_0 - \widehat{x})}{\|w\|}\right| = \left|\frac{w^T x_0}{\|w\|} - \frac{w^T \widehat{x}}{\|w\|}\right| \tag{66}$$

with the fact that $\omega^T \widehat{x} = -b$ we have the squared distance:

$$\|x_0 - \widehat{x}\|^2 = \left|\frac{w^T x_0 + b}{\|w\|}\right|^2. \tag{67}$$

Problem A.10

a.

$$f(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} y_i B_{ij} x_j + c = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i A_{ij} x_j + y_i B_{ij} x_j) + c \tag{68}$$

b.

Summations over indices:

$$\frac{\partial f(x,y)}{\partial x_k} = \sum_{i=1, i\neq k}^{n} A_{ik} x_i + \sum_{j=1, j\neq k}^{n} A_{kj} x_j + \sum_{i=1}^{n} y_i B_{ik} + 2A_{kk} x_k = \sum_{i=1}^{n} (A_{ik} x_i + A_{ki} x_i + y_i B_{ik}) \tag{69}$$

Vector notation:

$$\nabla_x f(x,y) = A^T x + Ax + B^T y \tag{70}$$

c.

Summations over indices:

$$\frac{\partial f(x,y)}{\partial y_k} = \sum_{i=1}^{n} B_{ki} x_i. \tag{71}$$

Vector notation:

$$\nabla_y f(x,y) = Bx. \tag{72}$$

Problem A.11

a. + b.

```
(base) cycyang@chengyesmbp2020 [22:20:40] [~/Coursework/CSE546/hw0]
-> % python p11.py
Problem A.11.a
[[ 0.125 -0.625  0.75 ]
 [-0.25   0.75  -0.5  ]
 [ 0.375 -0.375  0.25 ]]
Problem A.11.b
[[-2.]
 [ 1.]
 [-1.]]
[[6]
 [8]
 [7]]
```

Figure 1: The result of $A^{-1}, A^{-1}b$ and Ac using NumPy.

```python
import numpy as np

A = np.array(
    [
        [0, 2, 4],
        [2, 4, 2],
        [3, 3, 1]
    ]
)
A_inverse = np.linalg.inv(A)

print('Problem A.11.a')
print(A_inverse)

b = np.array(
    [
        [-2], [-2], [-4]
    ]
)
c = np.array(
    [
        [1], [1], [1]
    ]
)

print('Problem A.11.b')
print(np.dot(A_inverse, b))
print(np.dot(A, c))
```

Figure 2: The code screenshot from Problem A.11.

Problem A.12

a.

For all $x \in \mathcal{R}$:

$$\sqrt{\mathbb{E}[[](\widehat{F}_n(x) - F(x))^2]} \leq 0.0025 = \frac{1}{\sqrt{4n}} \tag{73}$$

therefore we choose $n = 40000$.

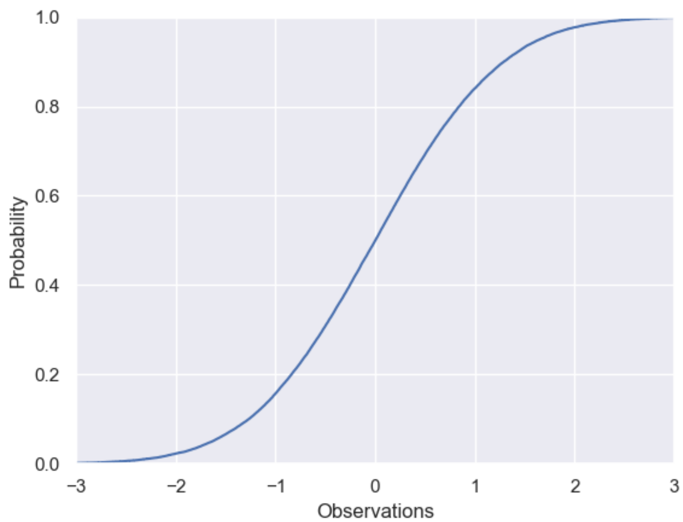


Figure 3: Plot of $\widehat{F}_n(x)$ from -3 to 3.

b.

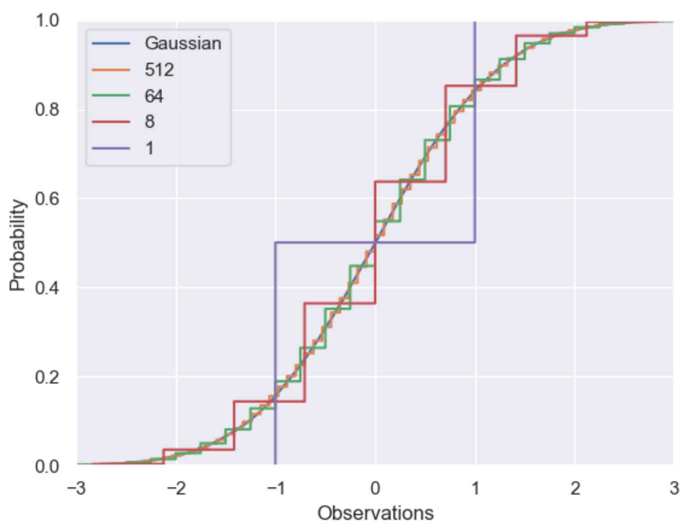


Figure 4: Plot of CDFs of $Y^{(}k)$.

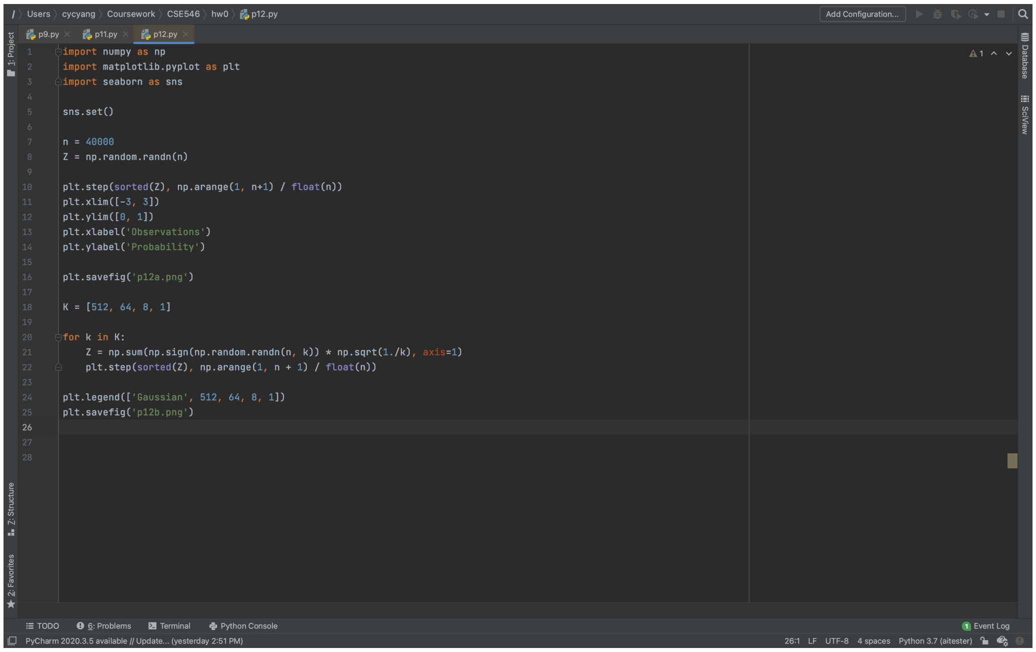


Figure 5: The code screenshot from Problem A.12.

Problem B.1

Since $X_1, \cdots, X_n$ are independent and identically distributed random variables, we have:

$$F_Y(y) = P(Y \leq y) = P(max\{X_1, \cdots, X_n\} \leq y) = \prod_{i=1}^{n} P(X_i \leq y) = (F_X(y))^n \tag{74}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = ny^{n-1} \tag{75}$$

then the expectation of Y can be derived as:

$$E[Y] = \int_0^1 y \cdot (ny^{n-1})dy = n \int_0^1 y^n = \frac{n}{n+1}. \tag{76}$$

Problem B.2

For any $x > 0$, by using Markov's inequality:

$$\begin{aligned}
P(X \geq \mu + \sigma x) &= P(x - \mu \geq \sigma x) &(77)\\
&\leq P(|x - \mu| \geq \sigma x) &(78)\\
&= P((x - \mu)^2 \geq \sigma^2 x^2) &(79)\\
&\leq \frac{E[(x - \mu)^2]}{\sigma^2 x^2} &(80)\\
&= \frac{\sigma^2}{\sigma^2 x^2} &(81)\\
&= \frac{1}{x^2} &(82)
\end{aligned}$$

Problem B.3

$$Tr(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}B_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n} B_{ji}\sum_{i=1}^{n} = \sum_{j=1}^{m}(BA)_{jj} = Tr(BA) \tag{83}$$

Problem B.4