

CSE 446 HW0

Zachary McNulty

April 2019

1 Problem 1

Let D represent the event that you have the disease and P be the event that you test positive for the disease. This question is then concerned with the probability $P(D | P)$. From the question, we see that:

$$P(D) = \frac{1}{10000}$$

$$P(P) = P(P | D)P(D) + P(P | D^C)P(D^C) = \frac{0.99}{10000} + \frac{0.01 * 9999}{10000} = \frac{100.98}{10000} = 0.010098$$

$$P(P | D) = 0.99$$

From here we can simply apply Bayes rule:

$$P(D | P) = \frac{P(P | D)P(D)}{P(P)} = \frac{0.99 * 0.0001}{0.010098} = 0.0098 \approx 1\%$$

Thus there is only around a 1% chance a randomly chosen person that tests positive actually has the disease.

2 Problem 2

a)

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - E[X]Y - E[Y]X + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Assuming that X and Y exist on the same probability space (which must be true as $E[Y | X = x] = x$ so Y can take on all values that X can) we can use the law of total expectation. By the law of total expectation we see the following:

$$E[Y] = E_X[E_{Y|X}[Y | X]] = E[X]$$

$$E[XY] = E_X[E_{XY|X}[XY | X]] = E_X[X * E_{Y|X}[Y | X]] = E[X^2]$$

Plugging these into the above equation yields

$$\begin{aligned} E[XY] - E[X]E[Y] &= E[X^2] - E[X]^2 \\ &= E[(X - E[X])^2] \end{aligned}$$

b)

In the above problem we showed that:

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] \\ E[XY] &= E_X[X * E_{Y|X}[Y | X]] \end{aligned}$$

Since X and Y are independent it holds that $E[Y | X] = E[Y]$. Thus:

$$E[XY] = E_X[X * E_{Y|X}[Y | X]] = E_X[XE[Y]] = E[X]E[Y]$$

Thus:

$$Cov(X, Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$$

3 Problem 3

a)

Let X, Y, Z be discrete random variables. Let f, g, h be their respective density functions and F, G, H their respective cumulative distribution functions. First, consider the cumulative distribution function $H(z) = P(Z \leq z)$. Since $Z = X + Y$ we can see:

$$\begin{aligned} P(Z \leq z) &= P(X + Y \leq z) \\ &= P(Y \leq z - X) \\ &= \sum_{x=-\infty}^{\infty} P(Y \leq z - x \mid X = x)P(X = x) \end{aligned}$$

Since X, Y are independent it holds that $P(Y \mid X) = P(Y)$. Thus:

$$\begin{aligned} &= \sum_{x=-\infty}^{\infty} P(Y \leq z - x)P(X = x) \\ H(z) &= \sum_{x=-\infty}^{\infty} G(z - x)f(x) \end{aligned}$$

Since $h(z) = \frac{dH(z)}{dz}$ all we have to do now is differentiate.

$$\begin{aligned} h(z) &= \frac{d}{dz} \sum_{x=-\infty}^{\infty} G(z - x)f(x) \\ &= \sum_{x=-\infty}^{\infty} g(z - x)f(x) \end{aligned}$$

This expression is analogous to the continuous case as even though in that case $P(X = x) = 0$ for any x , $f(x)$ is still a relative measure of the likelihood of a given x value and it can easily be verified that $h(z)$ is still a valid probability density function.

b)

$$\begin{aligned} h(z) &= \int_{-\infty}^{\infty} g(z - x)f(x)dx \\ &= \int_0^1 g(z - x)dx \\ h(z) &= \begin{cases} z & 0 \leq z \leq 1 \\ 2 - z & 1 < z \leq 2 \\ 0 & otherwise \end{cases} \end{aligned}$$

4 Problem 4

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ then we know $Y \sim N(a\mu + b, a^2\sigma^2)$. If we want $Y \sim N(0, 1)$ then this yields the system:

$$a\mu + b = 0$$

$$a^2\sigma^2 = 1$$

Solving this system yields $a = \frac{1}{\sigma^2}$ and $b = -\frac{\mu}{\sigma^2}$

5 Problem 5

$$\begin{aligned}
E[\sqrt{n}(\hat{\mu}_n - \mu)] &= \sqrt{n}E[\hat{\mu}_n - \mu] \\
&= \sqrt{n}(E[\hat{\mu}_n] - \mu) \\
&= \sqrt{n}(E[\frac{1}{n} \sum_{i=1}^n X_i] - \mu) \\
&= \sqrt{n}(\frac{1}{n}(E[X_1] + \dots + E[X_n]) - \mu) \\
&= \sqrt{n}(\frac{1}{n}(n\mu) - \mu) \\
&= 0
\end{aligned}$$

Note that $Var[X] = E[X^2] - E[X]^2$ and $Var[aX + b] = a^2Var[X]$. Thus:

$$\begin{aligned}
Var[\sqrt{n}(\hat{\mu}_n - \mu)] &= nVar[\hat{\mu}_n - \mu] \\
&= nVar[\hat{\mu}_n] \\
&= nE[(\frac{1}{n} \sum_{i=1}^n X_i - E[\hat{\mu}_n])^2] \\
&= \frac{1}{n}E[(\sum_{i=1}^n X_i - nE[\hat{\mu}_n])^2] \\
&= \frac{1}{n}E[(\sum_{i=1}^n X_i - n\mu)^2] \\
&= \frac{1}{n}E[(\sum_{i=1}^n (X_i - \mu))^2] \\
&= \frac{1}{n}E[\sum_{i,j \in \{1,2,\dots,n\}} (X_i - \mu)(X_j - \mu)] \\
&= \frac{1}{n} \sum_{i,j \in \{1,2,\dots,n\}} E[(X_i - \mu)(X_j - \mu)]
\end{aligned}$$

Since the X_i are iid random variables they are all pairwise independent. Thus:

$$Cov(X_i, X_j) = \begin{cases} \sigma^2 & i = j \\ 0 & otherwise \end{cases}$$

Thus it must hold that:

$$\frac{1}{n} \sum_{i,j \in \{1,2,\dots,n\}}^n E[(X_i - \mu)(X_j - \mu)] = \frac{1}{n} * n\sigma^2 = \sigma^2$$

Thus the mean and variance of $\sqrt{n}(\hat{\mu}_n - \mu)$ respectively are:

$$\begin{aligned} E[\sqrt{n}(\hat{\mu}_n - \mu)] &= 0 \\ \text{Var}[\sqrt{n}(\hat{\mu}_n - \mu)] &= \sigma^2 \end{aligned}$$

6 Problem 6

First we will calculate the expected value and variance of $1\{X \leq x\}$ with respect to x . We can think of this random variable as the function with a fixed x :

$$g(X) = \begin{cases} 1 & X \leq x \\ 0 & \text{otherwise} \end{cases}$$

Thus:

$$E[1\{X \leq x\}] = E[g(X)] = \int_{-\infty}^{\infty} g(s)f(s)ds = \int_{-\infty}^x f(s)ds = F(x)$$

Similarly, we can calculate the variance

$$Var[1\{X \leq x\}] = Var[g(X)] = E[g(X)^2] - E[g(X)]^2$$

Note that $g(X)^2 = g(X)$ for any value of X . Thus:

$$= E[g(X)] - E[g(X)]^2 = F(x) - F(x)^2 = F(x)(1 - F(x))$$

a)

$$E[\hat{F}_n(x)] = E\left[\frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}\right] = \frac{1}{n} (E[1\{X_1 \leq x\}] + E[1\{X_2 \leq x\}] + \dots + E[1\{X_n \leq x\}])$$

As we showed above $E[1\{X \leq x\}] = F(x)$ and since all the X_i are independent this holds true for all of them. Thus:

$$= \frac{1}{n} (n * F(x)) = F(x)$$

$$E[\hat{F}_n(x)] = F(x)$$

b)

As we showed in problem 5, the variance of $\frac{1}{n} \sum_{i=1}^n X_i$ for some sequence of iid random variables X_i is $\frac{Var[X]}{n}$. In this problem our random variable is $1\{X \leq x\}$ and we showed earlier it has variance $F(x)(1 - F(x))$. Thus:

$$Var[\hat{F}_n(x)] = \frac{F(x)(1 - F(x))}{n}$$

c)

$$E[(\hat{F}_n(x) - F(x))^2] = \text{Var}[\hat{F}_n(x)] = \frac{F(x)(1 - F(x))}{n}$$

As $F(x)$ is a cumulative distribution function, $F(x) \in [0, 1]$ for all x . We can find the maximum of $\frac{F(x)(1-F(x))}{n}$ using simple calculus:

$$\frac{d}{dF(x)} \frac{F(x)(1 - F(x))}{n} = \frac{1 - 2F(x)}{n} = 0$$

Solving for $F(x)$ yields a critical point at $F(x) = 0.5$. Taking the second derivative yields:

$$\frac{d^2}{dF(x)^2} \frac{F(x)(1 - F(x))}{n} = \frac{-2}{n}$$

Since $\frac{-2}{n} < 0$ for all natural numbers n , this function is concave down and thus $F(x) = 0.5$ is a maximum. Thus:

$$\frac{F(x)(1 - F(x))}{n} \leq \frac{0.5(1 - 0.5)}{n} = \frac{1}{4n}$$

Therefore:

$$E[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$$

7 Problem 7

a)

Calculating the reduced row echelon form for the first matrix yields:

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

We can see there are two nonzero rows so this matrix has rank 2.

Calculating the reduced row echelon form for the second matrix yields:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

We can see there are two nonzero rows so this matrix has rank 2.

b)

The rank of the matrix is the same as the minimum size of a basis for the column space so both matrices have a column space basis of size 2.

8 Problem 8

a)

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

b)

Define the augmented matrix $A \mid b$ as below:

$$\begin{bmatrix} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 3 & 3 & 1 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 3 & 3 & 1 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & -3 & -2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 4 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\text{Thus } x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

9 Problem 9

a)

The hyperplane $w^T x + b = 0$ with $w = [-1, 2]^T$ and $b = 2$ is just the line $-1x_1 + 2x_2 + 2 = 0$ in 2D-space with normal vector w .

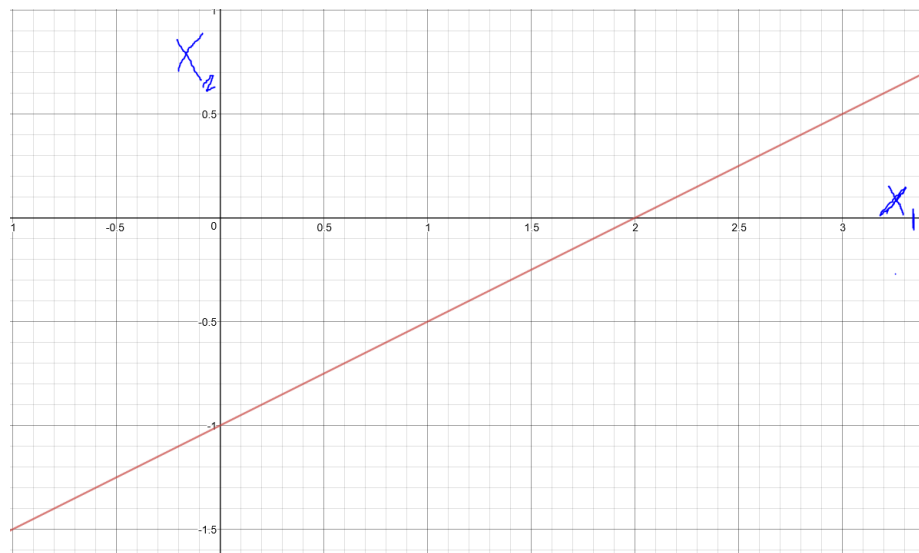


Figure 1: The hyperplane $[-1, 2]x + 2 = 0$ is a line in $2D$ space with normal vector $[-1, 2]$. The b term simply shifts the hyperplane $-\frac{b}{w_1}$ units in x_1 direction or equivalently $-\frac{b}{w_2}$ in the x_2 direction

b)

The hyperplane $w^T x + b = 0$ with $w = [1, 1, 1]^T$ and $b = 0$ is just a plane $x + y + z = 0$ in 3D-space: a plane with normal vector w that goes through the origin.

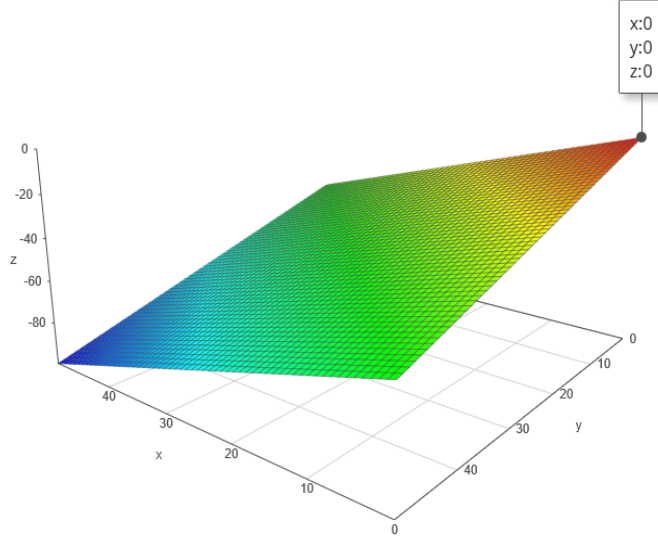


Figure 2: The hyperplane $[1, 1, 1][x, y, z]^T = 0$ is a plane in 3D space with normal vector $[1, 1, 1]$ that goes through the origin.

c)

Note that if \bar{x} is the point on the hyperplane $w^T x + b = 0$ that is nearest to the given point x_0 . We know that $w^T \bar{x} + b = 0 \rightarrow w^T \bar{x} = -b$. We can find \bar{x} by noting that the shortest path from the hyperplane to x_0 must be perpendicular to the hyperplane and thus in the direction of the hyperplane's normal vector w . Therefore, we know $\|x_0 - \bar{x}\|$ must simply be the magnitude of the component (scalar projection) of the vector $x_0 - \bar{x}$ in the direction of w which is $\frac{w \cdot (x_0 - \bar{x})}{\|w\|}$

$$\|x_0 - \bar{x}\| = \left| \frac{w^T (x_0 - \bar{x})}{\|w\|} \right| = \left| \frac{w^T x_0 + b}{\|w\|} \right|$$

Thus the squared distance to the hyperplane from x_0 is:

$$\|x_0 - \bar{x}\|^2 = \left(\frac{w^T x_0 + b}{\|w\|} \right)^2$$

10 Problem 10

a)

Note for a generic matrix $C \in \mathbb{R}^{n \times n}$ and vector $z \in \mathbb{R}^n$ where C_{*j} denotes the j th column of C and z_j the j th element of z , we can see that:

$$Cz = \sum_{j=1}^n C_{*j} z_j$$

This generates a column vector in \mathbb{R}^n . For two generic vectors $a, b \in \mathbb{R}^n$ the product $a^T b \in \mathbb{R}$ is:

$$a^T b = \sum_{i=1}^n a_i b_i$$

Overall then $a^T C b$ is:

$$a^T C b = \sum_{i=1}^n \sum_{j=1}^n a_i * C_{ij} b_j$$

This allows us to rewrite $f(x, y)$ as:

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

b)

$$\nabla_x f(x, y) = \left[\frac{\delta f(x, y)}{\delta x_1}, \frac{\delta f(x, y)}{\delta x_2}, \dots, \frac{\delta f(x, y)}{\delta x_k}, \dots, \frac{\delta f(x, y)}{\delta x_n} \right]^T$$

When calculating $\frac{\delta f(x, y)}{\delta x_k}$ we have to be careful to handle the case when $i = j = k$ since this generates a quadratic term $A_{kk} x_k^2$ rather than a simple linear term like all the others. This is what the final three terms of the below expression account for, even though they end up canceling.

$$\begin{aligned} \frac{\delta f(x, y)}{\delta x_k} &= \sum_{i=1}^n (y_i B_{ik} + x_i A_{ik}) + \sum_{j=1}^n (A_{kj} x_j) - x_k A_{kk} - A_{kk} x_k + 2x_k A_{kk} \\ &= \sum_{i=1}^n (y_i B_{ik} + x_i A_{ik} + A_{ki} x_i) \end{aligned}$$

c)

$$\nabla_x f(x, y) = \left[\frac{\delta f(x, y)}{\delta y_1}, \frac{\delta f(x, y)}{\delta y_2}, \dots, \frac{\delta f(x, y)}{\delta y_k}, \dots, \frac{\delta f(x, y)}{\delta y_n} \right]^T$$
$$\frac{\delta f(x, y)}{\delta y_k} = \sum_{j=1}^n B_{kj} x_j$$

11 Problem 11

```
A-1 :  
[[ 0.125 -0.625  0.75 ]  
 [-0.25  0.75  -0.5  ]  
 [ 0.375 -0.375  0.25 ]]  
  
A-1 b:  
[-2.  1. -1.]  
  
Ac :  
[6 8 7]
```

Figure 3: Results for calculating A^{-1} , $A^{-1}b$, and Ac

```
1  import numpy as np  
2  import matplotlib.pyplot as plt  
3  
4  
5  # Defining variables  
6  A = np.array([[0,2,4], [2,4,2], [3,3,1]])  
7  b = np.array([-2, -2, -4])  
8  c = np.array([1, 1, 1])  
9  
10 print("\nA:\n", A)  
11 print("\nb: \n", b)  
12 print("\nc: \n", c)  
13  
14  
15 # Problem 11  
16  
17 # part a)  
18 Ainv = np.linalg.inv(A)  
19  
20 print("\n A-1 : \n", Ainv)  
21 print("\nA * A-1: \n", np.dot(A, Ainv))  
22  
23 # part b)  
24  
25 Ainvb = np.dot(Ainv,b)  
26 Ac = np.dot(A,c)  
27  
28 print("\n A-1 b: \n", Ainvb)  
29 print("\n Ac : \n", Ac)  
30
```

Figure 4: Code that calculates A^{-1} , $A^{-1}b$, and Ac

12 Problem 12

a) and b)

We showed in problem 6c) that: $E[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$ so if we want $\sqrt{E[(\hat{F}_n(x) - F(x))^2]} \leq 0.0025 = \frac{1}{400}$ then we simply need to choose

$$\frac{1}{4n} \leq \frac{1}{400^2} \rightarrow n \geq 40000$$

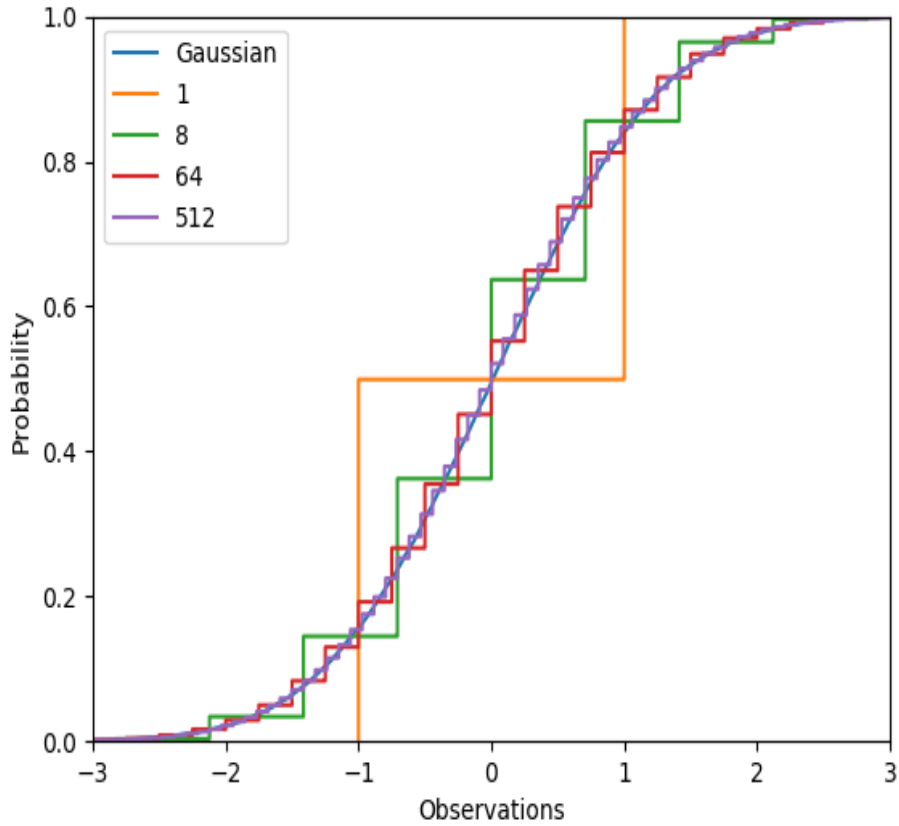


Figure 5: a plot of the empirical CDFs taken from a normal distribution


```

34 # Problem 12
35
36 # part a)
37 n = 40000
38 Z = np.random.randn(n)
39 plt.step(sorted(Z), np.arange(1, n+1)/float(n))
40
41 # part b)
42 for k in [1, 8, 64, 512]:
43     Yk = np.sum(np.sign(np.random.randn(n,k)) * np.sqrt(1./k), axis=-1)
44     plt.step(sorted(Yk), np.arange(1, n+1)/float(n))
45
46
47 # plotting results
48 plt.legend(['Gaussian', '1', '8', '64', '512'])
49 plt.ylim((0,1))
50 plt.xlim((-3,3))
51 plt.ylabel('Probability')
52 plt.xlabel('Observations')
53 plt.ioff()
54 plt.show()
55
56

```

Figure 6: Relevant code for this problem