CSE546 Assignment 0
A Problems

Joseph David

April 5 2021

1 Probability and Statistics

A.1 Letting A be the event that you have the disease and B that you test positive, we have that $P(B|A) = .99$ and $P(A) = .00001$. We use Bayes Rule and the fact that $P(B) = P(B \cap A) \cdot P(B \cap A^c) = (P(A) \cdot P(A|B)) + (P(A^c) \cdot P(A^c|B))$ to see that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{(.99)(.00001)}{(.0001)(.99) + (.9999)(.01)} = 1/102.$$

A.2(a) By the linearity of expectation, we have that $Cov(X,Y) = E[(X - E[X)(Y - E[Y])] = E[XY] - E[X]E[Y]$. Thus, the right hand side is just $Cov(X,X)$ so we may prove the problem if we can show that $E[XY] = E[X^2]$ and $E[Y] = E[X]$. For the first equality we have

$$\begin{aligned}
E[XY] &= \sum_z P(XY = z)z \\
&= \sum_z \sum_x P(X = x \cap Y = z/x)z \\
&= \sum_z \sum_x P(X = x)P(Y = z/x|X = x)z \\
&= \sum_x P(X = x) \sum_z P(xY = z|X = x)z \\
&= \sum_x P(X = x)E[xY|X = x] \\
&= \sum_x P(X = x)x^2 \ \text{ (by linearity of expectation and our assumption)} \\
&= E[X^2].
\end{aligned}$$

For the second equality, we have

$$\begin{aligned}
E[Y] &= \sum_y P(Y = y)y \\
&= \sum_y \sum_x P(X = x)P(Y = y|X = x)y \\
&= \sum_x P(X = x)E[Y|X = x] \\
&= \sum_x P(X = x)x \\
&= E[X].
\end{aligned}$$

This proves the result.

A.2(b) By part (a), if we show that $E[XY] = E[X]E[Y]$, then this will show that $Cov(X,Y) = 0$. To this end,

$$\begin{aligned}
E[X,Y] &= \sum_z P(XY = z)z \\
&= \sum_z \sum_x P(X = x \cap Y = z/x)z \\
&= \sum_z \sum_x P(X = x)P(xY = z)z \\
&= \sum_x P(X = x) \sum_z P(xY = z)z \\
&= \sum_x P(X = x)xE[Y] \\
&= E[X]E[Y].
\end{aligned}$$

A.3(a) By definition, $h(z) = d/dzP(X + Y \leq z)$. Let G be the cdf of Y. Then, we have that

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

A.3(b) Note that $h(z)$ is the length of the intersections of the supports of $f(x)$ and $g(z-x)$. Since f and g both have support in $[0,1]$, $h(z) = 0$ when $|z| \geq 1$. When $|z| < 1$, the intersection has length $1 - |z|$, so $h(z) = |1 - z|$ when $|z| < 1$. Combined, we see that

$$h(z) = \begin{cases} 1 - |z| & |z| \leq 1 \\ 0 & |z| > 1 \end{cases}$$

A.4 First, we wish that

$$0 = E[Y] = E[aX + b] = a\mu + b.$$

We also want the variance to equal 1, and therefore,

$$1 = E[Y^2] = a^2\sigma^2 + 2ab\mu + b^2.$$

Solving for a in the first equation gives $a = -b/\mu$ and plugging this into the second equation gives

$$\frac{b^2\sigma^2}{\mu^2} - b^2 = 1$$

and solving for b gives $b = \sqrt{\frac{1}{\frac{\sigma^2}{\mu^2} - 1}}$.

A.5 For the mean, we have

$$\begin{aligned}
E[\sqrt{n}(\widehat{\mu_n} - \mu)] &= \sqrt{n}E[\widehat{\mu_n} - \mu] \\
&= \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} E[X_i] - \mu) \\
&= \sqrt{n}(\mu - \mu) \\
&= 0.
\end{aligned}$$

For the variance, we first compute

$$\begin{aligned}
E[(\sqrt{n}(\widehat{\mu_n} - \mu) - E[\sqrt{n}(\widehat{\mu_n} - \mu)])^2] &= E[\sqrt{n}(\widehat{\mu_n} - \mu)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E[(X_i - \mu)(X_j - \mu)]].
\end{aligned}$$

However, by assumption $X_1, \ldots, X_n$ are independent random variables, and therefore for all $i \neq j$

$$E[(X_i - \mu)(X_j - \mu)] = Cov(X_i - \mu, X_j - \mu) = Cov(X_i, X_j) = 0,$$

by problem A.2. Thus, continuing our calculation,

$$\begin{aligned}
E[(\sqrt{n}(\widehat{\mu_n} - \mu) - E[\sqrt{n}(\widehat{\mu_n} - \mu)])^2] &= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \mu)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} Var(X_i - \mu, X_i - \mu) \\
&= \frac{1}{n}\sum_{i=1}^{n} Var(X_i, X_i) \\
&= \frac{1}{n} \cdot n\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

A.6(a) By linearity of expectation,

$$\begin{aligned}
E[\widehat{F_n}(x)] &= \frac{1}{n}\sum_{i=1}^{n} E[1\{X_i \leq x\}] \\
&= \frac{1}{n}\sum_{i=1}^{n} F(x) \\
&= F(x).
\end{aligned}$$

A.6(b) We compute the variance as

$$\begin{aligned}
Var[\widehat{F_n}] &= E[(\widehat{F_n}(x) - F(x))^2] \\
&= \frac{1}{n} E[(\sqrt{n}(\widehat{F_n}(x) - F(x)))^2] \\
&= \frac{1}{n} Var[1\{X_1 \leq x\}] \text{ (by A.6)} \\
&= \frac{1}{n}\left(E[1\{X_1 \leq x\}^2] - E[1\{X_1 \leq x\}]^2\right) \\
&= \frac{1}{n}\left(F(x) - F(x)^2\right) \\
&= \frac{F(x)(1 - F(x)}{n}.
\end{aligned}$$

A.6(c) By part (b), we have that

$$E[(\widehat{F_n}(x) - F(x))^2] = \frac{F(x)(1 - F(x))}{n}.$$

Thus, we need to find an upper bound on the right hand side. Consider the function $f(t) = t(1-t)$. We see that $f'(t) = 1 - 2t$ and $f''(t) = -2$, and hence from calculus we know that f has a maximum at $t = 1/2$ of value $f(1/2) = 1/4$. Therefore $F(x)(1 - F(x)) \leq 1/4$ which implies that

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

2 Linear Algebra and Vector Calculus

A.7(a) Both A and B have rank 2 because the third columns are linearly dependent on the first two columns, while the first two columns are linearly independent. For A, $3 \cdot \text{column1} - \text{column2} = \text{column3}$ and for B, $\text{column1} + \text{column2} = \text{column3}$. Since the rank is equal to the column span, they have rank 2.

A.7(b) By part (a), the column space of A and B each have dimension 2 and thus the minimal size basis for each column span is 2.

A.8(a)

$$Ac = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \\ 7 \end{pmatrix}.$$

A.8(b) We have the system of equations

$$2y + 4z = -2 \tag{1}$$

$$2x + 4y + 2z = -2 \tag{2}$$

$$3x + 3y + z = -4 \tag{3}$$

Solving (1) for z, we obtain

$$z = -.5 - .5y,$$

and if we plug this into (2) and (3) we obtain

$$3x + 2.5y = -3.5$$

$$2x + 3y = -1.$$

Solving this system gives $x = -2$ and $y = 1$, and after plugging y in (1), we see that $z = -1$. Thus, our final solution is

$$\mathbf{x} = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}.$$

A.9(a) Using the given definition of a hyperplane, we see that this hyperplane (a line in $\mathbb{R}^2$) is defined by

$$-x + 2y = -2.$$

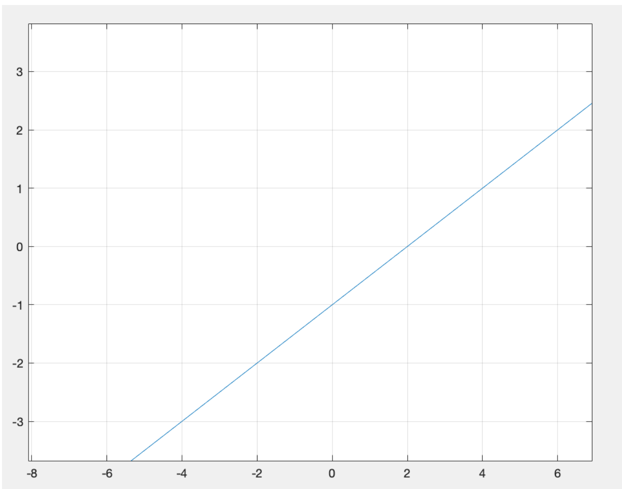

Figure 1: A.9(a): Line defined by $-x + 2y = -2$, plotted in Matlab.

A.9(b) This hyperplane is defined by

$$x + y + z = 0.$$

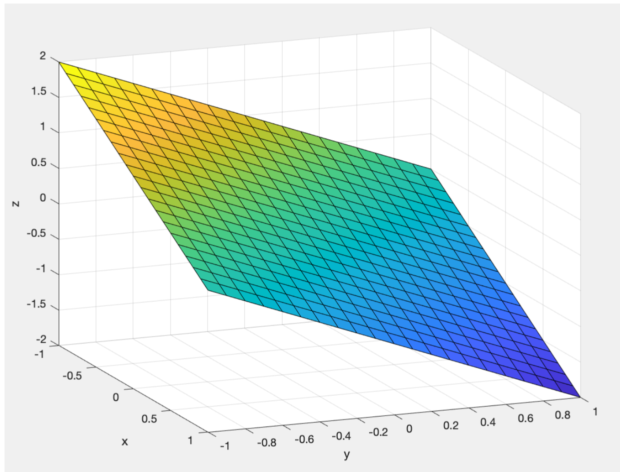

Figure 2: A.9(b): Plane defined by $x + y + z = 0$, plotted in Matlab.

A.9(c) Let x_0, w, and b be as in the problem. We may solve the simplified problem where we translate the plane P and x_0 by $-b$ so that P passes through the origin (ie, $b = 0$). Since we are translating, the distance between the new point and plane will be the same as the original, so this will give us the same result. Now, the distance from x_0 to P is simply the magnitude of the projection of x_0 onto the normalized normal vector $w/||w||_2$, so the distance is given by

$$\left| \frac{w^T x_0}{||w||_2} \right|.$$

Thus, the squared distance is

$$\left| \frac{w^T x_0}{||w||_2} \right|^2.$$

A.10(a)

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

A.10(b) First, we compute for $k \in \{1, \ldots, n\}$,

$$\begin{aligned} \frac{\partial f}{\partial x_k}(x, y) &= \frac{\partial}{\partial x_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i [A_{ij} x_j + B_{ij} y_j]) + c \right) \\ &= 2A_{kk} x_k + B_{kk} y_k + \sum_{j \neq k} [A_{kj} x_j + B_{kj} y_j] + \sum_{i \neq k} [x_i A_{ik}]. \end{aligned}$$

Therefore,

$$
\begin{aligned}
\nabla_x f(x,y) &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x,y) e_k \\
&= \sum_{k=1}^{n} \left(2A_{kk}x_k + B_{kk}y_k + \sum_{j\neq k}[A_{kj}x_j + B_{kj}y_j] + \sum_{i \neq k}[x_i A_{ik}] \right) e_k,
\end{aligned}
$$

where e_k is the kth standard basis vector.

A.10(c) For $k \in \{1, \ldots, n\}$,

$$
\begin{aligned}
\frac{\partial f}{\partial y_k}(x,y) &= \frac{\partial}{\partial y_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i[A_{ij}x_j + B_{ij}y_j]) + c \right) \\
&= \sum_{i=1}^{n} x_i B_{ik},
\end{aligned}
$$

so we have

$$
\begin{aligned}
\nabla_y f(x,y) &= \sum_{k=1}^{n} \frac{\partial f}{\partial y_k}(x,y) e_k \\
&= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} x_i B_{ik} \right) e_k.
\end{aligned}
$$

3 Programming

A.11(a),(b)

```
In [18]: Ainv = inv(A)
         Ainv

Out[18]: matrix([[ 0.125, -0.625,  0.75 ],
                 [-0.25 ,  0.75 , -0.5  ],
                 [ 0.375, -0.375,  0.25 ]])

In [19]: Ainv*b

Out[19]: matrix([[-2.],
                 [ 1.],
                 [-1.]])

In [20]: A*c

Out[20]: matrix([[6],
                 [8],
                 [7]])
```

Figure 3: A.11(a),(b): Results for A^{-1}, $A^{-1}b$, and Ac.

A.12(a),(b) By A.6, we solve for n and find $n = 40,000$.

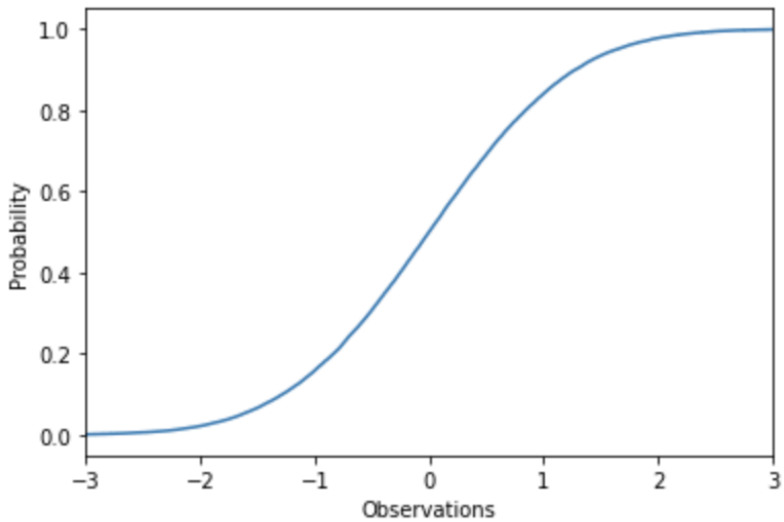

Figure 4: Results for A.12a

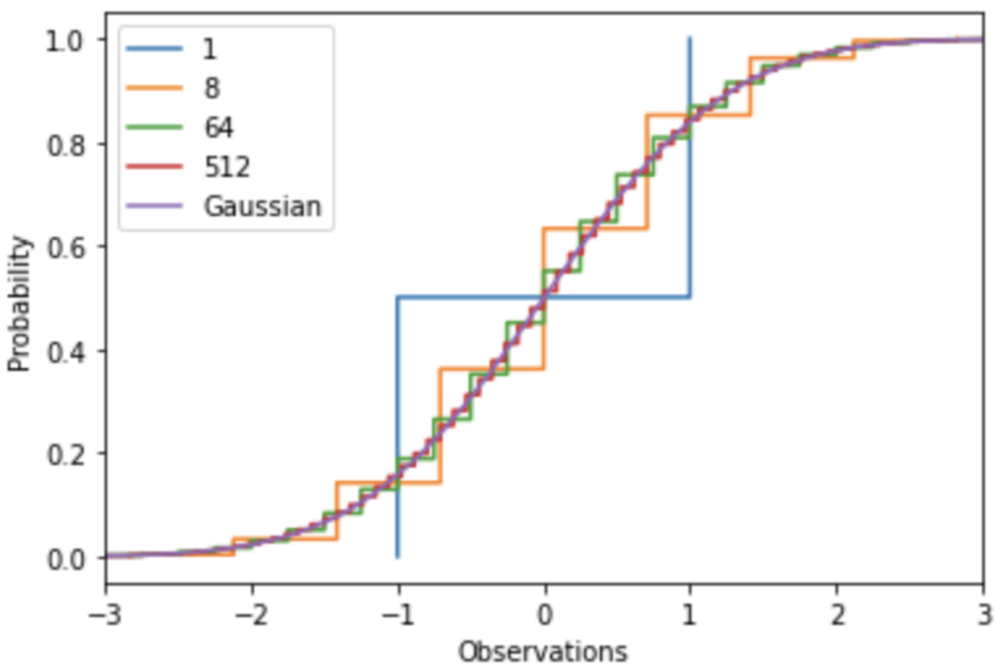

Figure 5: Results for A.12b

```python
K = [1, 8, 64, 512]
for k in K:
    Y = np.sum(np.sign(np.random.randn(40000,k))*np.sqrt(1./k), axis=1)
    plt.step(sorted(Y), np.arange(1,40001)/float(40000), label = k)
Z = np.random.randn(40000)
plt.step(sorted(Z), np.arange(1,40001)/float(40000), label = 'Gaussian')
plt.xlabel('Observations')
plt.ylabel('Probability')
plt.xlim([-3,3])
plt.legend()
```

Figure 6: A.12 : Code for producing graph.