STAT 184, PROBLEM SET 0

Matthew Qu
Due: September 8, 2022
matthewqu@college.harvard.edu

1. (**Policies**)

Proof. Collaborators: Kevin Huang
Acknowledgements: Problem 4.1: STAT 110 textbook
I have read these policies. □

2. (**Certify that you have read the website**)

Proof. I have read the course policies on the website. $\square$

3. (Bayes' Rule)

Proof. Let D be the event that you have the disease, and let T be the event that you test positive. We want to find the conditional probability $\mathbb{P}(D \mid T)$. By Bayes' Rule and LOTP, we have

$$\begin{aligned}\mathbb{P}(D \mid T) &= \frac{\mathbb{P}(T \mid D)\mathbb{P}(D)}{\mathbb{P}(T \mid D)\mathbb{P}(D) + \mathbb{P}(T \mid D^c)\mathbb{P}(D^c)} \\ &= \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999} \\ &= \frac{1}{102}.\end{aligned}$$

The probability that you have the disease is less than 1%, or about 0.0098. $\square$

4. (Probability)

Proof.

1. Let $H(z)$ denote the CDF of Z. We can condition on X and apply LOTP to see that

$$\begin{aligned} H(z) &= \mathbb{P}(X + Y \leq z) \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X + Y \leq z \mid X = x) f(x)\, dx \\ &= \int_{-\infty}^{\infty} \mathbb{P}(Y \leq z - x \mid X = x) f(x)\, dx. \end{aligned}$$

Now, since Y is independent of X, the probability inside the integral is simply $\mathbb{P}(Y \leq z - x)$. Now, to find the density of Z, we differentiate with respect to z, which we can evaluate by taking the derivative of the integrand. Since $\frac{d}{dz}\mathbb{P}(Y \leq z - x) = g(z - x)$, it follows that

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

2.

(a) Using part 1, we know that $h(z) = \int_{-\infty}^{\infty} f(x)g(z - x)\, dx$. Both f and g are 1 on $[0, 1]$ and 0 otherwise; in particular, the integrand will be 1 if $0 \leq x \leq 1$ and $0 \leq z - x \leq 1$. If $0 < z < 1$, then the integrand is 1 for $x \in [0, z]$. Therefore, we have $h(z) = \int_0^z 1\, dx = z$ on this interval. Similarly, if $1 < z < 2$, then the integrand is 1 for $x \in [z - 1, 1]$, and so we have $h(z) = \int_{1-z}^{1} 1\, dx = 2 - z$. For all other values of z, the integrand is 0. Therefore, we have the piecewise density

$$h(z) = \begin{cases} z & \text{if } 0 \leq z \leq 1 \\ 2 - z & \text{if } 1 \leq z \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

We can check that this density does indeed integrate to 1.

(b) Consider the geometric argument. The line $X + Y = \frac{5}{4}$ intersects the unit square at $(\frac{1}{4}, 1)$ and $(1, \frac{1}{4})$. The area corresponding to the event $X + Y \geq \frac{5}{4}$ is the upper triangle above this line with area $\frac{9}{32}$. Conditioned on this, the event that $X \leq \frac{1}{2}$ is the upper left triangle with vertices $(\frac{1}{4}, 1), (\frac{1}{2}, 1)$, and $(\frac{1}{2}, \frac{3}{4})$. This triangle has area $\frac{1}{32}$. Since X and Y are Uniform, probabilities are proportional to area, so we have

$$\mathbb{P}\left(X \leq \frac{1}{2} \mid X + Y \geq \frac{5}{4}\right) = \frac{\frac{1}{32}}{\frac{9}{32}} = \frac{1}{9}.$$

3. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then we can standardize X by subtracting its mean and dividing by its standard deviation. That is, $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. It follows that $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$.

4.

(a) If $\mathbb{E}(Y \mid X = x) = x$, then $\mathbb{E}(Y \mid X) = X$. After taking the expectation of both sides, Adam's Law implies that $\mathbb{E}(Y) = \mathbb{E}(X)$. Now, let us use Adam's Law on the definition for covariance, conditioning on X:

$$\begin{aligned} \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) &= \mathbb{E}(\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \mid X) \\ &= \mathbb{E}((X - \mathbb{E}(X))\mathbb{E}(Y - \mathbb{E}(Y) \mid X)). \end{aligned}$$

Now, we note that $\mathbb{E}(Y - \mathbb{E}(Y) \mid X) = \mathbb{E}(Y \mid X) - \mathbb{E}(Y) = X - \mathbb{E}(X)$. It follows that the above expression simplifies to $\mathbb{E}((X - \mathbb{E}(X))^2)$. Thus, we have

$$\operatorname{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big),$$

as desired.

(b) Note that the definition of covariance can be rewritten as $\operatorname{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. If X and Y are independent with densities f and g, respectively, then their joint density is simply $f(x)g(y)$. It follows that

$$\begin{aligned}\mathbb{E}(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x)g(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} xf(x)\,dx \int_{-\infty}^{\infty} yg(y)\,dy \\ &= \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

Therefore, we conclude that $\operatorname{Cov}(X, Y) = 0$.

5.

(a) By linearity, we have

$$\begin{aligned}\mathbb{E}(\widehat{F}_n(x)) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\mathbf{1}\{X_i \le x\}) \\ &= \frac{1}{n}\sum_{i=1}^{n} F(x) \\ &= F(x).\end{aligned}$$

(b) We are given that $\operatorname{Var}(\widehat{F}_1(x)) = \mathbb{E}\big((\widehat{F}_1(x) - F(x))^2\big)$. Expanding, this yields

$$\operatorname{Var}(\widehat{F}_1(x)) = \mathbb{E}\big(\mathbf{1}\{X_1 \le x\}^2\big) - 2F(x)\mathbb{E}(\mathbf{1}\{X_1 \le x\}) + F(x)^2.$$

However, note that $I^a = I$ for any indicator random variable I. Therefore, $\mathbb{E}\big(\mathbf{1}\{X_1 \le x\}^2\big) = \mathbb{E}(\mathbf{1}\{X_1 \le x\}) = F(x)$, and it follows that

$$\operatorname{Var}(\widehat{F}_1(x)) = F(x) - 2F(x)^2 + F(x)^2 = F(x)(1 - F(x)).$$

(c) Because the X_i are iid, we have

$$\begin{aligned}\operatorname{Var}(\widehat{F}_n(x)) &= \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(\mathbf{1}\{X_i \le x\}) \\ &= \frac{1}{n}\operatorname{Var}(\widehat{F}_1(x)).\end{aligned}$$

From part (b), it follows that $\operatorname{Var}(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.

(d) For any $x \in \mathbb{R}$, note that $F(x)(1 - F(x))$ has a maximum value of $\frac{1}{4}$ when $F(x) = \frac{1}{2}$. In other words, $F(x)(1 - F(x)) \le \frac{1}{4}$ for all $x \in \mathbb{R}$. Thus, we have

$$\mathbb{E}\big((\widehat{F}_n(x) - F(x))^2\big) = \operatorname{Var}(\widehat{F}_n(x)) \le \frac{1}{4n}.$$

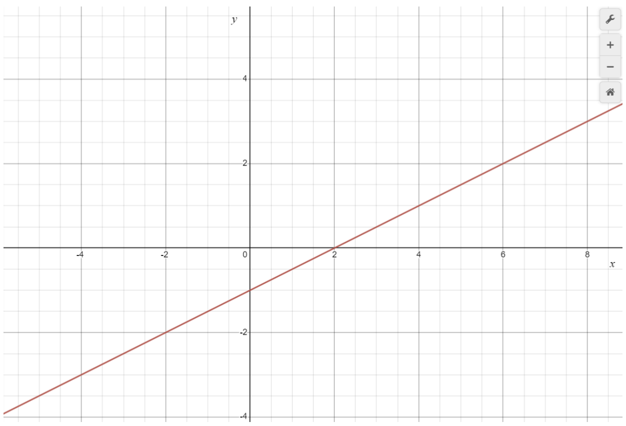
$\square$

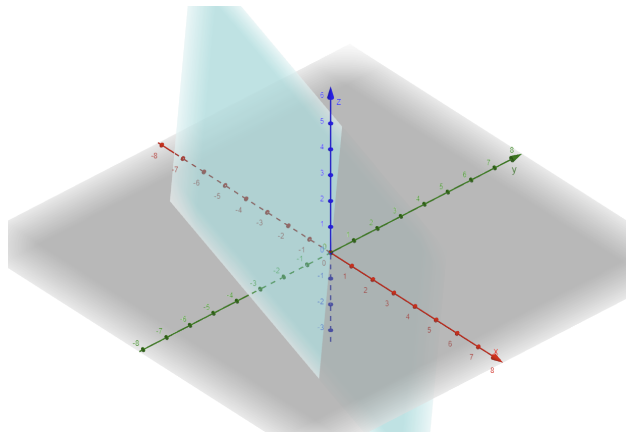
5. (Geometry and Linear Algebra)

Proof.

1.

(a) The labels x, y, and z correspond to the first, second, and third coordinates of the vector $\vec{x}$, respectively.





(b) See the right plot above.

(c) Let H denote the hyperplane defined by $w^\top x + b = 0$. We know that w is the normal vector to H. Therefore, if we subtract a multiple of w from x_0, we will get a vector x in H. In other words, there exists a constant c such that $x_0 - cw = x$, where x satisfies $w^\top x + b = 0$. Thus, we have $w^\top(x_0 - cw) = w^\top x$, which we can rewrite as $w^\top x_0 + c\|w\|^2 = -b$. Solving for c, we get

$$c = \frac{w^\top x_0 + b}{\|w\|^2}.$$

However, note that c is not the distance between x_0 and the hyperplane; rather, it is $\|cw\|$. Therefore, the distance is

$$\|cw\| = \frac{|w^\top x_0 + b|}{\|w\|},$$

which means that the squared distance is $\frac{(w^\top x_0 + b)^2}{\|w\|^2}$.

2.

(a) Let c_1, c_2, and c_3 denote the first, second, and third column vectors, respectively. We see that $3c_1 - c_2 = c_3$, so the third column is a linear combination of the first two. However, c_2 is not a scalar multiple of c_1, and so the first two columns are linearly independent. Therefore, the rank of A is 2.

(b) A simple minimal basis for the column span of A would be the first two column vectors (c_1, c_2).

3.

(a) The product Ac will simply be the row-wise sums of A, so $Ac = [6, 8, 7]^\top$.

(b) Let $x = [x_1, x_2, x_3]^\top$. We then have the equations

$$\begin{aligned} 2x_2 + 4x_3 &= -2 \\ 2x_1 + 4x_2 + 2x_3 &= -2 \\ 3x_1 + 3x_2 + x_3 &= -4. \end{aligned}$$

Using equations 2 and 3 to eliminate x_1 yields $6x_2 + 4x_3 = 2$. With equation 1, it follows that $x_2 = 1$ and $x_3 = -1$. This means that $x_1 = -2$. Thus, the solution to $Ax = b$ is $x = [-2, 1, -1]^\top$.

4. First, consider the product $\mathbf{A}x$. The i^{th} coordinate of $\mathbf{A}x$ is the sum $\sum_{j=1}^{n} A_{ij}x_j$. Therefore, it follows that

$$x^\top \mathbf{A}x = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij}x_j \right).$$

We can rewrite this as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j.$$

The second product $y^\top \mathbf{B}x$ is analogous. Therefore, we have

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

$\square$

6. (Programming)

Proof.

1.

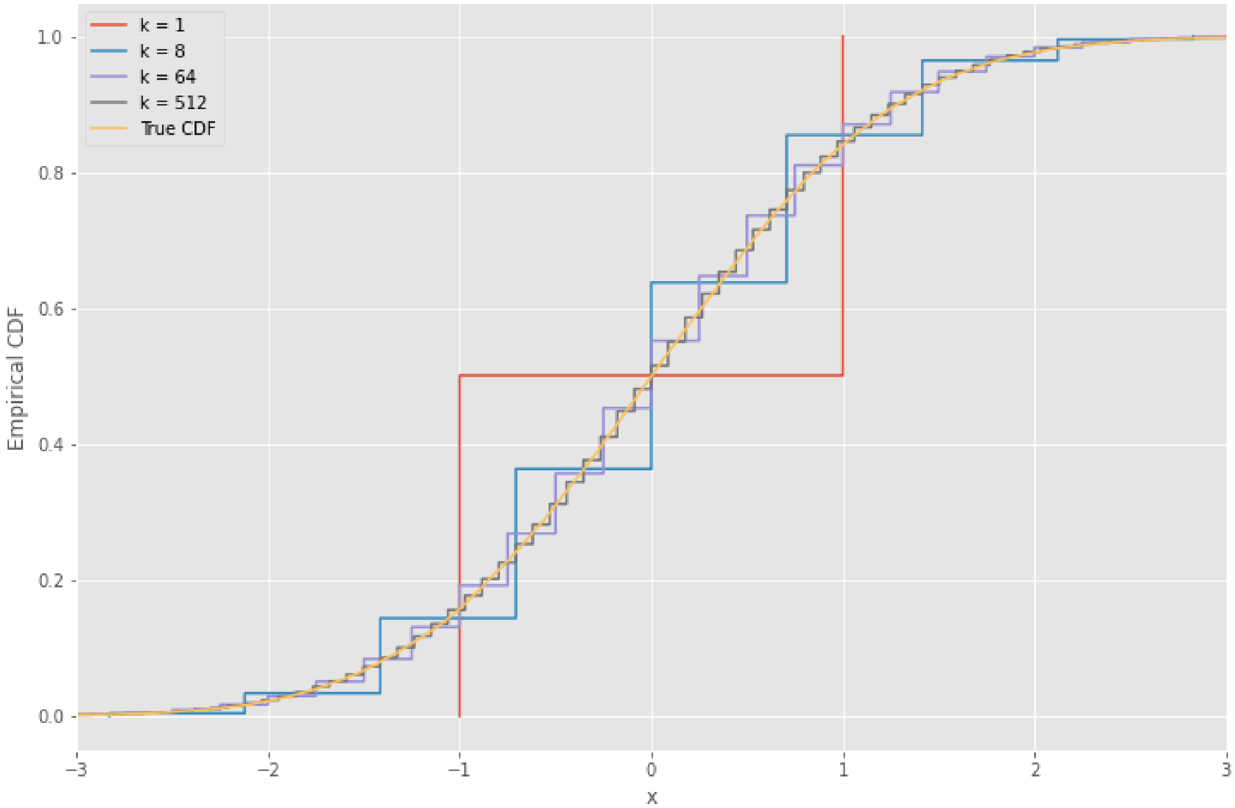
(a) We have

$$A^{-1} = \begin{bmatrix} 0.125 & -0.625 & 0.75 \\ -0.25 & 0.75 & -0.5 \\ 0.375 & -0.375 & 0.25 \end{bmatrix}$$

(b) $A^{-1}b$ and Ac are calculated below, and agree with the results from problem 5.3.

```
In [2]: A = np.array([[0, 2, 4],
                      [2, 4, 2],
                      [3, 3, 1]])
        A

Out[2]: array([[0, 2, 4],
               [2, 4, 2],
               [3, 3, 1]])

In [3]: A_inv = np.linalg.inv(A)
        A_inv

Out[3]: array([[ 0.125, -0.625,  0.75 ],
               [-0.25 ,  0.75 , -0.5  ],
               [ 0.375, -0.375,  0.25 ]])

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

In [5]: A_inv.dot(b)

Out[5]: array([-2.,  1., -1.])

In [6]: A.dot(c)

Out[6]: array([6, 8, 7])
```

2. The plot for each value of k along with the true CDF is shown below. From problem 4.5, we find that we should choose $n = 40000$ to ensure that the expectation is bounded correctly. We see that as k grows large, the empirical CDF approaches the true normal CDF, which is consistent with the CLT.


k = 1
k = 8
k = 64
k = 512
True CDF
1.0
0.8
0.6
0.4
0.2
0.0
Empirical CDF
−3
−2
−1
0
1
2
3
x

□