CSE446 HW0

Brian Kang

Probability and Statistics

A.1 *[2 points]* (Bayes Rule, from Murphy exercise 2.4.) After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you dont have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease? (Show your calculations as well as giving the final result.)

Denote negative test result as (-), positive as (+), not having the disease as (not have), having it as (have). We know that each respective pairs are mutually exclusive (you can't be both infected and not infected).
We are given that $P(-|nothave) = 0.99 = P(+|have)$ and $P(have) = 1/10000 = 0.0001$.
Then by Bayes Rule:

$$\begin{aligned} P(have|+) &= \frac{P(+|have) * P(have)}{P(+|have) * P(have) + P(+|nothave) * P(nothave)} \\ &= \frac{0.99 * 0.0001}{0.99 * 0.0001 + (1 - P(-|nothave)) * (1 - 0.0001)} \\ &= \frac{0.99 * 0.0001}{0.99 * 0.0001 + (1 - 0.99) * (0.9999)} \\ &= \frac{0.99 * 0.0001}{0.99 * 0.0001 + 0.01 * 0.9999} \\ &= \boxed{0.0098 = 0.98\%} \end{aligned} \tag{1}$$

A.2 For any two random variables X, Y the *covariance* is defined as $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. You may assume X and Y take on a discrete values if you find that is easier to work with.

a. *[1 points]* If $\mathbb{E}[Y|X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Proof.
First, let $Z = \mathbb{E}[Y|X]$, then $\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$ **(A)** and $\mathbb{E}[XZ] = \mathbb{E}[X\mathbb{E}[Y|X]] = \mathbb{E}[\mathbb{E}[XY|X]] = \mathbb{E}[XY]$ **(B)**.
Second, given $\mathbb{E}[Y|X = x] = x$, $\mathbb{E}[\mathbb{E}[Y|X]] = \sum_{\text{all x}} \mathbb{E}[Y|X = x] * P(X = x) =$
$= \sum_{\text{all x}} x * P(X = x) = \mathbb{E}[X]$. Hence, $\mathbb{E}[X] = \mathbb{E}[Y]$ with the given **(C)**.
(Sidenote, a related problem that we can now show is that $\text{Cov}(X, Y) = \text{Cov}(X, \mathbb{E}[Y|X])$)

Thus, given $\mathbb{E}[Y|X=x]=x$,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])*(Y-\mathbb{E}[Y])] \\
&= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY - X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X]] - \mathbb{E}[Y\mathbb{E}[X]] \text{ by (C)} \\
&= \mathbb{E}[X\mathbb{E}[Y|X] - X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X]] - \mathbb{E}[X]\mathbb{E}[Y] \text{ by (B)} \\
&= \mathbb{E}[X]\mathbb{E}[Y|X] + \mathbb{E}[-X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X]] - \mathbb{E}[X]\mathbb{E}[X] \text{ by (C)} \\
&= \mathbb{E}[XX - X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[X]] \text{ by (A) and (C)} \\
&= \mathbb{E}[XX - X\mathbb{E}[X] - X\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[X]] \\
&= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\
&= \mathbb{E}[(X-\mathbb{E}[X])^2]
\end{aligned} \tag{2}$$

□

b. *[1 points]* If X, Y are independent show that $\mathrm{Cov}(X,Y)=0$.

Proof.

If X, Y are independent,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])*(Y-\mathbb{E}[Y])] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= 0
\end{aligned} \tag{3}$$

□

A.3 Let X and Y be independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$.

a. *[2 points]* Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$. (If you are more comfortable with discrete probabilities, you can instead derive an analogous expression for the discrete case, and then you should give a one sentence explanation as to why your expression is analogous to the continuous case.).

Proof.

Let J be the joint of X and Y.
Then, since X and Y are independent, the joint density function of J=$f(x)g(x)$.
Then, (let us denote the integral of g as g^* as we proceed when appropriate)

$$\begin{aligned}
P(X+Y\leq z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x)g(y)dydx \\
&= \int_{-\infty}^{\infty} f(x)(\int_{-\infty}^{z-x} g(y)dy)dx \\
&= \int_{-\infty}^{\infty} f(x)(g^*(z-x) - g^*(-\infty))dx \\
&= \int_{-\infty}^{\infty} f(x)(g^*(z-x) - 0)dx
\end{aligned} \tag{4}$$

This is the CDF of Z. Differentiate once to get the PDF of Z, which is h.

$$\frac{d}{dz}g(z)^* = g \text{ so, } h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

$\square$

b. *[1 points]* If X and Y are both independent and uniformly distributed on $[0,1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0,1]$ and 0 otherwise) what is h, the PDF of $Z = X + Y$?

$f(x) = 1$ if $0 \le x \le 1$, so $h(z) = \int_0^1 g(z-x)dx$.
If $0 \le z \le 1$, then $h(z) = \int_0^z 1dx = \boxed{z}$.
If $z < 0$ or $z > 2$, then $h(z) = \boxed{0}$.
If $1 < z \le 2$, then $h(z) = \int_{z-1}^1 1dx = \boxed{2-z}$.

A.4 *[1 points]* A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2. Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0,1)$.

We know that $\frac{1}{\sigma}(X - \mu) \sim \mathcal{N}(0,1)$.
Then $\boxed{a = 1/\sigma \text{ and } b = -\mu/\sigma}$.

A.5 *[2 points]* For a random variable Z, its mean and variance are defined as $\mathbb{E}[Z]$ and $\mathbb{E}[(Z - \mathbb{E}[Z])^2]$, respectively. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, each with mean μ and variance σ^2. If we define $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$, what is the mean and variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$?

$$\mathbb{E}[\widehat{\mu}_n] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}n\mu = \mu$$

$$\mathbb{V}[\widehat{\mu}_n] = \mathbb{V}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n^2}\mathbb{V}[\sum_{i=1}^n X_i] = \frac{1}{n^2}(\mathbb{V}[X_1] + \mathbb{V}[X_2] + \cdots + \mathbb{V}[X_n]) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$\begin{aligned}
\mathbb{E}[\sqrt{n}(\widehat{\mu}_n - \mu)] &= \sqrt{n}\mathbb{E}[\mu - \mu] \\
&= \boxed{0} \\
\mathbb{V}[\sqrt{n}(\widehat{\mu}_n - \mu)] &= (\sqrt{n})^2\mathbb{V}[\widehat{\mu}_n - \mu] \\
&= n\mathbb{V}[\widehat{\mu}_n] \\
&= n\frac{\sigma^2}{n} \\
&= \boxed{\sigma^2}
\end{aligned} \tag{5}$$

A.6 If $f(x)$ is a PDF, the cumulative distribution function (CDF) is defined as $F(x) = \int_{-\infty}^x f(y)dy$. For any function $g : \mathbb{R} \to \mathbb{R}$ and random variable X with PDF $f(x)$, recall that the expected value of $g(X)$ is defined as $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(y)f(y)dy$. For a boolean event A, define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Thus, $\mathbf{1}\{x \le a\}$ is 1 whenever $x \le a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \le x\}]$. Let $X_1, \ldots, X_n$ be *independent and identically distributed* random variables with CDF $F(x)$. Define $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \le x\}$. Note, for every x, that $\widehat{F}_n(x)$ is an *empirical estimate* of $F(x)$. You may use your answers to the previous problem.

a. *[1 points]* For any x, what is $\mathbb{E}[\widehat{F}_n(x)]$?

$$\begin{aligned}\mathbb{E}[\widehat{F}_n(x)] &= \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \leq x\}] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathbf{1}\{X_i \leq x\}] \\ &= \frac{1}{n}\sum_{i=1}^{n}F(x) \\ &= \frac{1}{n}nF(x) \\ &= \boxed{F(x)}\end{aligned} \tag{6}$$

b. *[1 points]* For any x, the variance of $\widehat{F}_n(x)$ is $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]$. Show that Variance$(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.

Proof.
$\mathbf{1}\{X \leq x\}$ follows a Bernoulli distribution with $P(X = 1) = F(x) = P(X \leq x)$ and $P(X = 0) = 1 - P(X = 1) = P(X > x)$.
Then we know that

$$\begin{aligned}\mathbb{V}[\mathbf{1}\{X \leq x\}] &= \mathbb{E}[\mathbf{1}\{X \leq x\}^2] - E[\mathbf{1}\{X \leq x\}]^2 \\ &= (F(x) * 1^2 + (1 - F(x)) * 0^2) - F(x)^2 = F(x)(1 - F(x))\end{aligned}$$

Now, from iid X_i and problem A.5, we know that

$$\begin{aligned}\mathbb{V}[\widehat{F}_n(x)] &= \mathbb{V}[\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \leq x\}] \\ &= \frac{1}{n^2}n\mathbb{V}[\mathbf{1}\{X_i \leq x\}] \\ &= \frac{F(x)(1 - F(x))}{n}\end{aligned} \tag{7}$$

□

c. *[1 points]* Using your answer to b, show that for all $x \in \mathbb{R}$, we have $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$.

Proof.
We know that $0 \leq F(x) \leq 1$, implying $0 \leq (1 - F(x)) \leq 1$.
$F(x)(1 - F(x))$ is a quadratic with respect to $F(x)$ with the value of the quadratic maximized at $F(x) = 1/2$.
Then, $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2] = \dfrac{F(x)(1 - F(x))}{n} \leq \dfrac{\frac{1}{2} * \frac{1}{2}}{n} = \dfrac{\frac{1}{4}}{n} = \dfrac{1}{4n}$. □

Linear Algebra and Vector Calculus

A.7 (Rank) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. For each matrix A and B,

For A, $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$

For B, $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & -1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

a. *[2 points]* what is its rank?

Both A's and B's rank are $\boxed{2}$.

b. *[2 points]* what is a (minimal size) basis for its column span?

Both A's and B's basis are $\boxed{\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right)}$.

A.8 (Linear equations) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

a. *[1 points]* What is Ac?

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \boxed{\begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}}$$

b. *[2 points]* What is the solution to the linear system $Ax = b$? (Show your work).

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

$\vec{x} = \boxed{\begin{bmatrix} -2 & 1 & -1 \end{bmatrix}^T}$

A.9 (Hyperplanes) Assume w is an n-dimensional vector and b is a scalar. A hyperplane in $\mathbb{R}^n$ is the set $\{x : x \in \mathbb{R}^n, \text{ s.t. } w^T x + b = 0\}$.

a. *[1 points]* ($n = 2$ example) Draw the hyperplane for $w = [-1, 2]^T$, $b = 2$? Label your axes.

Graph is in next page.

b. *[1 points]* ($n = 3$ example) Draw the hyperplane for $w = [1, 1, 1]^T$, $b = 0$? Label your axes.

Graph is in next page.

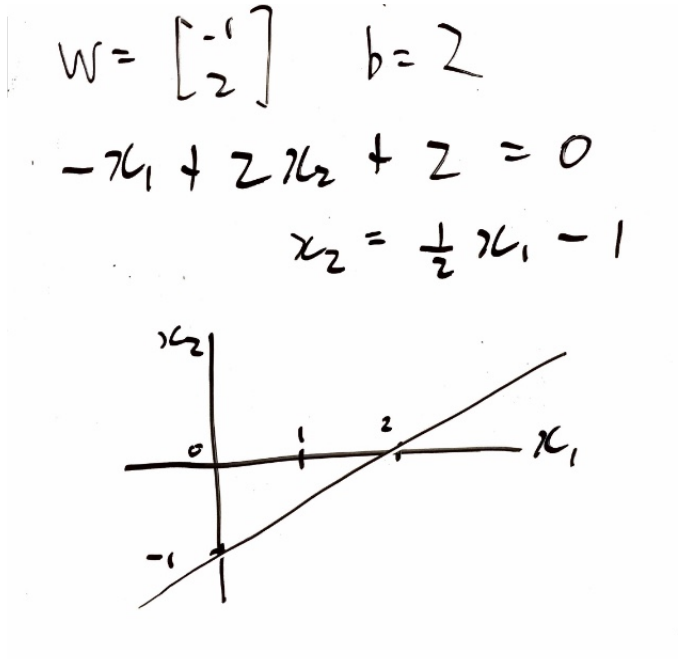


Figure 1: A.9.a Figure

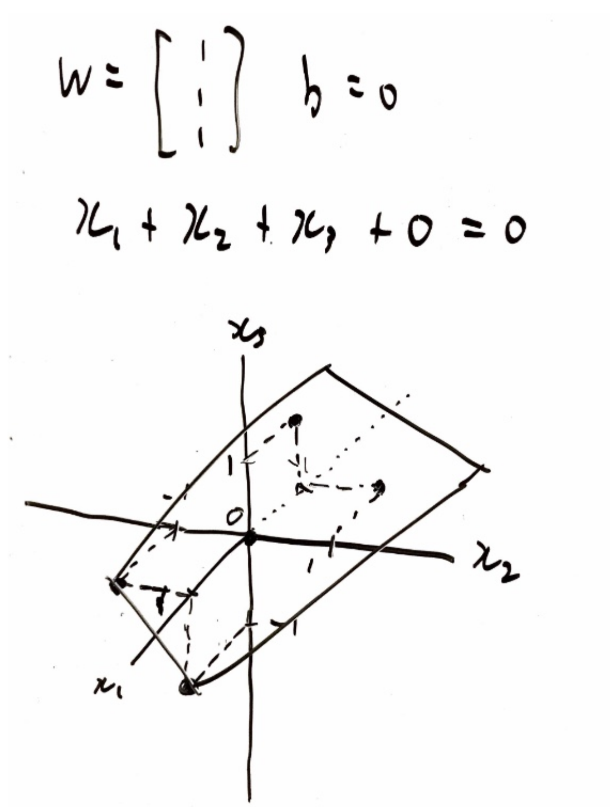


Figure 2: A.9.b Figure

c. *[2 points]* Given some $x_0 \in \mathbb{R}^n$, find the *squared distance* to the hyperplane defined by $w^T x + b = 0$. In other words, solve the following optimization problem:

$$\min_x \|x_0 - x\|^2$$
$$\text{s.t. } w^T x + b = 0$$

(Hint: if $\widetilde{x}_0$ is the minimizer of the above problem, note that $\|x_0 - \widetilde{x}_0\| = |\frac{w^T(x_0 - \widetilde{x}_0)}{\|w\|}|$. What is $w^T \widetilde{x}_0$?)

The equation of our plane is given by $\vec{w}^T \vec{x} + b = 0$ with normal vector $\vec{w}$. With some point $\vec{x_0}$, the vector from the plane to the point is given by $\vec{u} = -(\vec{x} - \vec{x_0})$. The minimum distance from the point to plane is the projection $proj_{\vec{w}} \vec{u}$. We want to solve the *minimum squared distance*, which is $(proj_{\vec{w}} \vec{u})^2$. Then,

$$\begin{aligned}
(D)^2 &= (|proj_{\vec{w}} \vec{u}|)^2 \\
&= \left(\frac{|\vec{w}^T \vec{u}|}{||\vec{w}||} \right)^2 \\
&= \left(\frac{|-b - \vec{w}^T \vec{x_0}|}{||\vec{w}||} \right)^2 \\
&= \left(\frac{b + \vec{w}^T \vec{x_0}}{||\vec{w}||} \right)^2 \\
&= \boxed{\frac{\left(b + \vec{w}^T \vec{x_0}\right)}{\sum_i^n w_i^2}}
\end{aligned} \tag{8}$$

A.10 For possibly non-symmetric $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^T \boldsymbol{A} x + y^T \boldsymbol{B} x + c$. Define $\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial z_1} & \frac{\partial f(x,y)}{\partial z_2} & \dots & \frac{\partial f(x,y)}{\partial z_n} \end{bmatrix}^T$.

a. *[2 points]* Explicitly write out the function $f(x, y)$ in terms of the components $A_{i,j}$ and $B_{i,j}$ using appropriate summations over the indices.

We know $x^T \boldsymbol{A} x = x \cdot (\boldsymbol{A} x)$. Likewise, $y^T \boldsymbol{B} x = y \cdot (\boldsymbol{B} x)$.

$$f(x, y) = x^T \boldsymbol{A} x + y^T \boldsymbol{B} x + c = \sum_{i=1}^n (\sum_{j=1}^n a_{ij} x_j)(x^T)_i + \sum_{i=1}^n (\sum_{j=1}^n b_{ij} x_j) y_i + c$$

b. *[2 points]* What is $\nabla_x f(x, y)$ in terms of the summations over indices *and* vector notation?

Vector notation:

$$\begin{aligned}
\nabla_x f(x, y) &= \frac{\partial}{\partial x}(x^T \boldsymbol{A} x) + \frac{\partial}{\partial x}(y^T \boldsymbol{B} x) + 0 \\
&= \boldsymbol{A} x + \frac{\partial}{\partial x}(x^T \boldsymbol{A}^T x) + 0 + \frac{\partial}{\partial x}(x^T \boldsymbol{B}^T y) \\
&= \boldsymbol{A} x + \boldsymbol{A}^T x + \boldsymbol{B}^T y \\
&= (\boldsymbol{A} + \boldsymbol{A}^T) x + \boldsymbol{B}^T y
\end{aligned} \tag{9}$$

Summation notation extracting from vector notation:

$$
\begin{aligned}
\nabla_x f(x,y) &= \boldsymbol{A}x + \boldsymbol{A}^T x + \boldsymbol{B}^T y \\
&= \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j & \sum_{j=1}^n a_{2j}x_j & \dots & \sum_{j=1}^n a_{nj}x_j \end{bmatrix}^T \\
&+ \begin{bmatrix} \sum_{j=1}^n a_{j1}x_j & \sum_{j=1}^n a_{j2}x_j & \dots & \sum_{j=1}^n a_{jn}x_j \end{bmatrix}^T \\
&+ \begin{bmatrix} \sum_{j=1}^n b_{j1}y_j & \sum_{j=1}^n b_{j2}y_j & \dots & \sum_{j=1}^n b_{jn}y_j \end{bmatrix}^T
\end{aligned} \tag{10}
$$

c. *[2 points]* What is $\nabla_y f(x,y)$ in terms of the summations over indices *and* vector notation?

Vector notation:

$$
\begin{aligned}
\nabla_y f(x,y) &= \frac{\partial}{\partial y}(x^T \boldsymbol{A}x) + \frac{\partial}{\partial y}(y^T \boldsymbol{B}x) + 0 \\
&= 0 + \boldsymbol{B}x
\end{aligned} \tag{11}
$$

Summation notation extracting from vector notation:

$$
\begin{aligned}
\nabla_y f(x,y) &= \boldsymbol{B}x \\
&= \begin{bmatrix} \sum_{j=1}^n b_{1j}x_j & \sum_{j=1}^n b_{2j}x_j & \dots & \sum_{j=1}^n b_{nj}x_j \end{bmatrix}^T
\end{aligned} \tag{12}
$$

Programming

A.11 For the A, b, c as defined in Problem 8, use NumPy to compute (take a screen shot of your answer):

a. *[2 points]* What is A^{-1}?

b. *[1 points]* What is $A^{-1}b$? What is Ac?

A.11.a&b solutions below.

Python code for A.11

```
import numpy as np
#import matplotlib.pyplot as plt

A = np.array([[0,2,4],[2,4,2],[3,3,1]])
b = np.array([-2,-2,-4]).T  # transpose
c = np.array([1,1,1]).T

# A.11.a
Ainv = np.linalg.inv(A) # A inverse
print("A^(-1) = \n", Ainv)

# A.11.b
Ainvb = Ainv @ b  #A inverse * b
Ac = A @ c    #A * c
print("\nAb = \n", Ainvb)
print("\nAc = \n", Ac)
```

```
A^(-1) =
 [[ 0.125 -0.625  0.75 ]
 [-0.25   0.75  -0.5  ]
 [ 0.375 -0.375  0.25 ]]

Ab =
 [-2.  1. -1.]

Ac =
 [6 8 7]
```

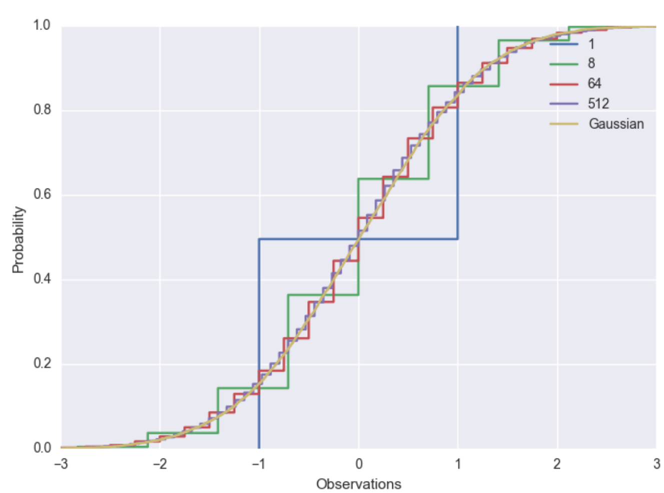
Figure 3: A.11.a&b Answer

A.12 *[4 points]* Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y, respectively, are equal, i.e. for all x, $|F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance-$1/k$ random variables converges to a (standard) Normal distribution as k goes off to infinity. We will study this phenomenon empirically (you will use the Python packages Numpy and Matplotlib). Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} B_i$ where each B_i is equal to -1 and 1 with equal probability. From your solution to problem 5, we know that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$.

a. For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. If $F(x)$ is the true CDF from which each Z_i is drawn (i.e., Gaussian) and $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x\}$, use the answer to problem 1.5 above to choose n large enough such that, for all $x \in \mathbb{R}$, $\sqrt{\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]} \leq 0.0025$, and plot $\widehat{F}_n(x)$ from -3 to 3.
(Hint: use `Z=numpy.random.randn(n)` to generate the random variables, and `import matplotlib.pyplot as plt`;
`plt.step(sorted(Z), np.arange(1,n+1)/float(n))` to plot).

b. For each $k \in \{1, 8, 64, 512\}$ generate n independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.
(Hint: `np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1./k), axis=1)` generates n of the $Y^{(k)}$ random variables.)

Be sure to always label your axes. Your plot should look something like the following (Tip: checkout `seaborn` for instantly better looking plots.)



Python code for A.12

```python
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt

sns.set()   # for the grid on the background

# From A.6, E[] <= 1/(4n)
# Now we want sqrt(E[]) <= 0.0025,
#    i.e. E[] <= 0.0025^2 = 1/(4n).
#    Thus, n = 40,000
n = 40000
Z = np.random.randn(n)
line1, = plt.step(sorted(Z), np.arange(1, n + 1) / float(n))
line1.set_label("Gaussian")

k = [1,8,64,512]
for i in k:
    v = np.sum(np.sign(np.random.randn(n, i))*np.sqrt(1./i), axis=1)
    line2, = plt.step(sorted(v), np.arange(1, n + 1) / float(n))
    line2.set_label(i)

plt.xlim(-3, 3)
plt.ylim(0.0,1.0)
plt.xlabel("Observations")
plt.ylabel("Probability")
plt.legend()
plt.show()
```

My plots below.

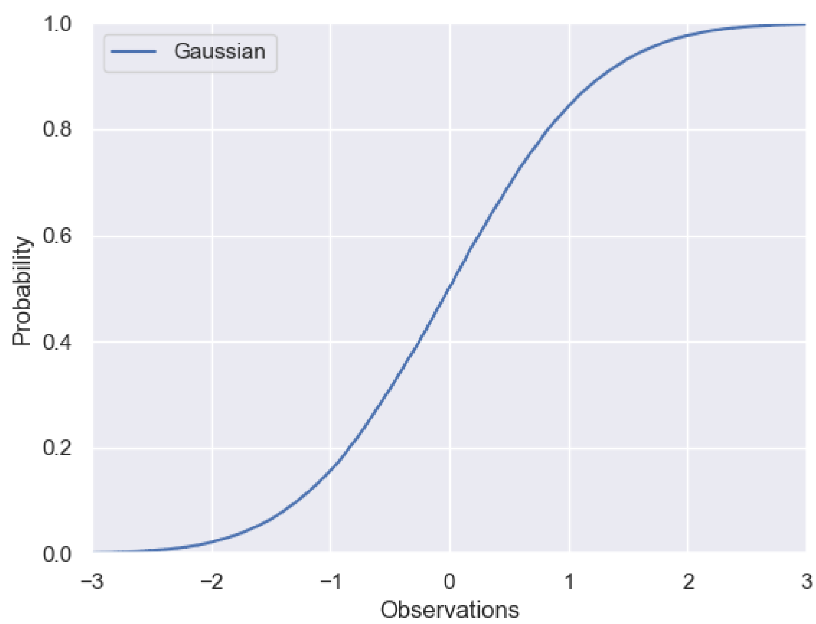


Figure 4: My A.12.a plot.

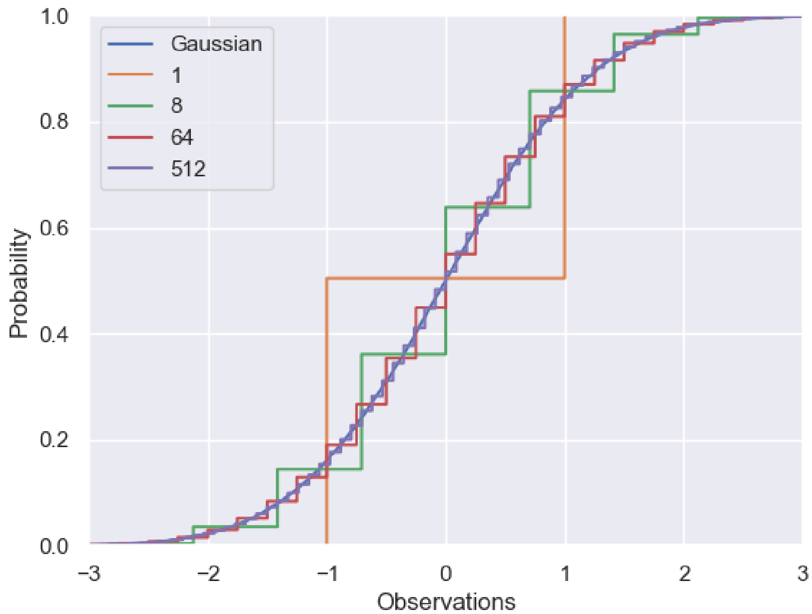


Figure 5: My A.12.b plot.