CSE 546 Homework #0

Lu Yu

Spring 2020

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.)

We want to find out that the probability that people actually having the disease.
P be the proposition that test is positive, P- be the proposition that test is negative.
D be the proposition of having disease, D- be the proposition of not having disease.
Bayes Rule: $P(D|P) = \frac{P(P|D)P(D)}{P(P)}$.
We already know:

$$P(P|D) = 0.99$$
$$P(P-|D-) = 0.99$$
$$P(D) = \frac{1}{10000} = 0.0001$$
$$P(D-) = 1 - P(D) = 0.9999$$
$$P(P|D-) = 1 - P(P|D) = 1 - 0.99 = 0.01$$

So,

$$P(P) = P(P|D-) \times P(D-) + P(P|D) \times P(D) = 0.01 \times 0.9999 + 0.99 \times 0.0001 = 0.010098$$

So,

$$P(P|D) = \frac{P(P|D)P(D)}{P(P)} = \frac{0.99 \times 0.0001}{0.010098} = 0.0098$$

A.2 For any two random variables X, Y the *covariance* is defined as $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. 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 $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

$$f(x,y) = f_y(y|x)f_x(x) = xf_x(x)$$
$$\int_{-\infty}^{\infty} yf_y(y|x)dy = \mathbb{E}(Y|X = x) = x$$

$$
\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dydx \\
&= \int_{-\infty}^{\infty} xf_x(x)(\int_{-\infty}^{\infty} yf_y(y|x)dy)dx \\
&= \int_{-\infty}^{\infty} xf_x(x)xdx \\
&= \mathbb{E}(X^2)
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx \\
&= \int_{-\infty}^{\infty} f_x(x)(\int_{-\infty}^{\infty} yf_y(y|x)dy)dx \\
&= \int_{-\infty}^{\infty} f_x(x)xdx \\
&= \mathbb{E}(X)
\end{aligned} \tag{2}
$$

$$
\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}(X^2) - \mathbb{E}(X)\mathbb{E}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]
$$

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

By independence, $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$, and $X - \mathbf{E}[X]$, $Y - \mathbf{E}[Y]$ are independent.

$$
Cov(X,Y) = \mathbf{E}(X - \mathbf{E}[X])(Y - \mathbf{E}[Y]) = \mathbf{E}(X - \mathbf{E}[X])\mathbf{E}(Y - \mathbf{E}[Y]) = (\mathbf{E}(X) - \mathbf{E}[X])(\mathbf{E}(Y) - \mathbf{E}[Y]) = 0
$$

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.).

$$
H(z) = P(Z) = P(X+Y) = \int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} f(x,y)dy]dx
$$

$$
h(z) = \frac{d}{dz}H(z) = \int_{-\infty}^{\infty}[\frac{d}{dz}\int_{-\infty}^{\infty} f(x,y)dy]dx = \int_{-\infty}^{\infty} f(x, z-x)dx
$$

Since X and Y are independent random variables, then

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

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 $Z = X + Y$?

$Z = X + Y$, so the range is [0,2].
From a., we know that $h(z) = \int_{-\infty}^{\infty} f(x, z-x)dx$, so

$$
f(x, z-x) = f(x) \times g(y) = \begin{cases} 1, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & otherwise \end{cases}
$$

Determine the integration range of x, $0 \le y \le 1 \to 0 \le z - x \le 1 \to z - 1 \le x \le z$ and $0 \le x \le 1$
(1) $z \le 0, h(z) = 0$

(2) $\begin{cases} z-1<0 \\ z\geq 0 \end{cases} \rightarrow 0 \leq z < 1$
So, the integration range of x is from 0 to z.

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{0}^{z} 1dx = z$$

(3)$z-1 \geq 0 \rightarrow z \geq 1$
So, the integration range of x is from z-1 to z, and $z \geq 1$, therefore, the integration range of x is from z-1 to 1 .

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{z-1}^{1} 1dx = 2-z$$

$h(z) = 2-z \geq 0 \rightarrow z \leq 2$, so $1 \leq z \leq 2$
So, the PDF of Z is:

$$h(z) = \begin{cases} z, & 0 \leq z < 1 \\ 2-z, & 1 \leq z \leq 2 \\ 0, & otherwise \end{cases}$$

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 $Y \sim \mathcal{N}(0, 1)$.

$$X \sim \mathcal{N}(\mu,\, \sigma^2)$$
$$Y = aX + b \sim \mathcal{N}(a\mu + b,\, a^2\sigma^2) = \mathcal{N}(0,1)$$
$$\begin{cases} a^2\sigma^2 = 1 \\ a\mu + b = 0 \end{cases}$$
$$\begin{cases} a = \pm\sqrt{\frac{1}{\sigma^2}} \\ b = \mp\mu\sqrt{\frac{1}{\sigma^2}} \end{cases}$$

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 $X_1, \dots, X_n$ 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 $\sqrt{n}(\widehat{\mu}_n - \mu)$?

$$\mathbb{E}(\widehat{\mu_n}) = \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum \mu = \mu$$
$$\mathbb{E}(\sqrt{n}(\widehat{\mu}_n - \mu)) = \sqrt{n}(\mathbb{E}(\widehat{\mu}_n) - \mathbb{E}(\mu)) = 0$$
$$D(\widehat{\mu_n}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\sum \sigma^2 = \frac{\sigma^2}{n}$$
$$D(\sqrt{n}(\widehat{\mu}_n - \mu)) = nD(\widehat{\mu}_n - \mu) = nD(\widehat{\mu}_n) + D(\mu) - 2Cov(\widehat{\mu}_n, \mu)$$

Since μ is a constant,

$$D(\sqrt{n}(\widehat{\mu}_n - \mu)) = nD(\widehat{\mu}_n) = \sigma^2$$

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 $g : \mathbb{R} \rightarrow \mathbb{R}$ 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, $\mathbf{1}\{x \leq a\}$ is 1 whenever $x \leq a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]$. Let $X_1, \dots, X_n$ 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 $\mathbb{E}[\widehat{F}_n(x)]$?
From A.5, $\mathbb{E}[\widehat{F}_n(x)] = \mathbb{E}[\mathbf{1}\{X \le x\}] = F(x)$. So, for any x, $\mathbb{E}[\widehat{F}_n(x)] = F(x)$

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$(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.
Since $\mathbf{1}\{x \le a\}$ is 1 whenever $x \le a$ and 0 whenever $x > a$, $\mathbb{E}[\widehat{F}_n(x)] = \mathbb{E}[\mathbf{1}\{X \le x\}] = \mathbb{E}[\mathbf{1}\{X\}]$, $D[\widehat{F}_n(x)] = D[\mathbf{1}\{X \le x\}] = D[\mathbf{1}\{X\}]$.
For boolean events, $\mathbb{E}[\mathbf{1}\{X > x\}] = 1 - \mathbb{E}[\mathbf{1}\{X \le x\}] = 1 - F(x)$, and $D(\mathbf{1}\{X\}) = nP(1-P) = n\mathbb{E}[\mathbf{1}\{X > x\}] \times \mathbb{E}[\mathbf{1}\{X \le x\}] = nF(x)(1-F(x))$.
From A.5, $D(\widehat{\mu_n}) = \frac{\sigma^2}{n}$. So, Variance$(\widehat{F}_n(x)) = \frac{D(\mathbf{1}\{X\})}{n} = \frac{nF(x)(1-F(x))}{n} = \frac{F(x)(1-F(x))}{n}$

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

$$\text{Variance}(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n} = \frac{-F^2(x) + F(x)}{n} = \frac{-(F(x) - \frac{1}{2})^2 + \frac{1}{4}}{n} \le \frac{1}{4n}$$

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,

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

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{r_3 - r_1} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{r_2 - r_1} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{2r_3 - r_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

So $rank(A) = 2$.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{r_3 - r_1} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix} \xrightarrow{r_2 - r_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & -1 & -1 \end{bmatrix} \xrightarrow{2r_3 - r_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

So $rank(B) = 2$.

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

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

$Col(A) = \mathbb{R}^2$and observe that first column in A is one-third sum of 2nd and 3rd columns.
So the basis for A's column span is

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

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

$Col(B) = \mathbb{R}^2$and observe that first column in A is difference of 2nd and 3rd columns. So the basis for B's column span is

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

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?

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

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

$$\begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix}$$

$$\begin{cases} 2x_2 + 4x_3 = -2 \\ 2x_1 + 4x_2 + 2x_3 = -2 \\ 3x_1 + 3x_2 + x_3 = -4 \end{cases}$$

$$\begin{bmatrix} x_1 = -2 \\ x_2 = 1 \\ x_3 = -1 \end{bmatrix} \rightarrow \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

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.

$$[-1, 2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2 = 0 \rightarrow -x_1 + 2x_2 + 2 = 0$$

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

$$[1, 1, 1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \rightarrow x_1 + x_2 + x_3 = 0$$

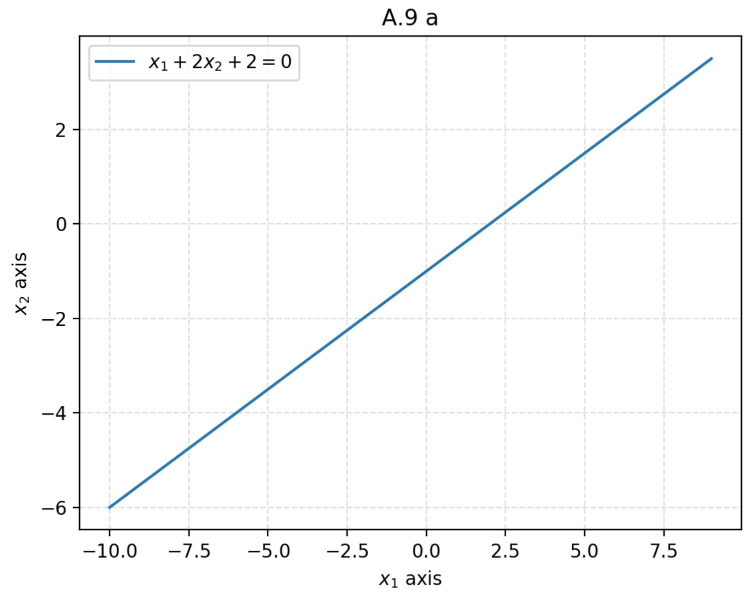

Figure 1: A.9 a

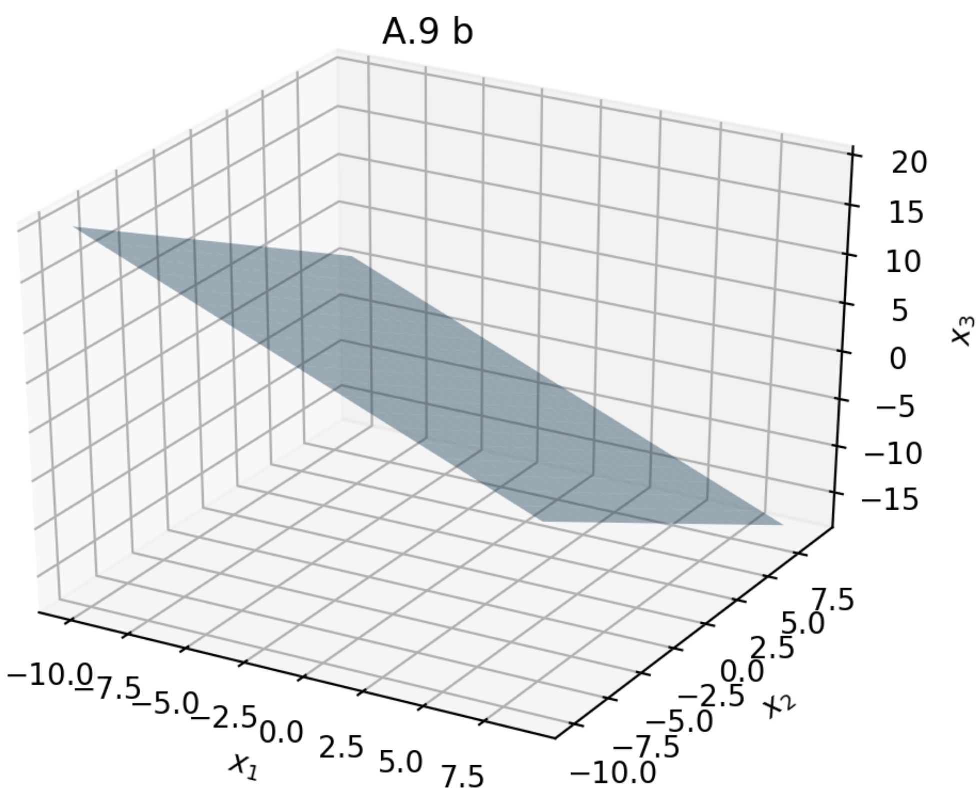

Figure 2: A.9 b

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 $w^T\widetilde{x}_0$?)

$$min||x_0 - x||^2$$
$$s.t. w^T x + b = 0$$

Given that $\tilde{x_0}$ is the solution of the above problem, so

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

$$||x_0 - x||^2 = \left|\frac{w^T x_0 + b}{||w||}\right|^2 = \frac{(w^T x_0 + b)^2}{w'w}$$

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.

$$x^T = [x_1, x_2, \dots, x_n], y^T = [y_1, y_2, \dots, y_n]$$

$$\begin{aligned} f(x,y) &= x^T \boldsymbol{A} x + y^T \boldsymbol{B} x + c \\ &= [\sum_{i=1}^n x_i A_{i,1}, \sum_{i=1}^n x_i A_{i,2}, \dots, \sum_{i=1}^n x_i A_{i,n}]x + [\sum_{i=1}^n y_i B_{i,1}, \sum_{i=1}^n y_i B_{i,2}, \dots, \sum_{i=1}^n y_i B_{i,n}]x + c \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i x_j A_{i,j} + \sum_{i=1}^n \sum_{j=1}^n y_i x_j B_{i,j} + c \end{aligned} \tag{3}$$

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

$\nabla_x f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x_1} & \frac{\partial f(x,y)}{\partial x_2} & \dots & \frac{\partial f(x,y)}{\partial x_n} \end{bmatrix}^T = \begin{bmatrix} \sum_{i=1}^n x_i A_{i,1} + \sum_{j=1}^n x_j A_{1,j} + \sum_{i=1}^n y_i B_{i,1} \\ \sum_{i=1}^n x_j A_{i,1} + \sum_{j=1}^n x_j A_{2,j} + \sum_{i=1}^n y_i B_{i,2} \\ \dots \\ \sum_{i=1}^n x_j A_{i,n} + \sum_{j=1}^n x_j A_{n,j} + \sum_{i=1}^n y_i B_{i,n} \end{bmatrix} = \boldsymbol{A} x + \boldsymbol{A}^T x + \boldsymbol{B}^T y$

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

$$\nabla_y f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial y_1} & \frac{\partial f(x,y)}{\partial y_2} & \dots & \frac{\partial f(x,y)}{\partial y_n} \end{bmatrix}^T = \begin{bmatrix} \sum_{j=1}^n x_j B_{1,j} & \sum_{j=1}^n x_j B_{2,j} & \dots & \sum_{j=1}^n x_j B_{n,j} \end{bmatrix}^T = \boldsymbol{B} x$$

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?

```
In [1]: import numpy as np
        A = np.array([[0, 2, 4],
            [2, 4, 2],
            [3, 3, 1]])
        b = np.array([[-2, -2, -4]]).T
        c = np.array([[1, 1, 1]]).T

In [2]: np.linalg.inv(A)

Out[2]: array([[ 0.125, -0.625,  0.75 ],
               [-0.25 ,  0.75 , -0.5  ],
               [ 0.375, -0.375,  0.25 ]])

In [3]: np.dot(np.linalg.inv(A),b)

Out[3]: array([[-2.],
               [ 1.],
               [-1.]])

In [4]: np.dot(A,c)

Out[4]: array([[6],
               [8],
               [7]])
```

Figure 3: A.11

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 $i = 1, \dots, n$ 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).
From A.6a, for any x, $\mathbb{E}[\widehat{F}_n(x)] = F(x)$, so we plot $F(x)$ instead of $\widehat{F}_n(x)$.
From A.6c, $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$, So $\frac{1}{4n} \leq 0.0025^2$, $n \geq 40000$.

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.)

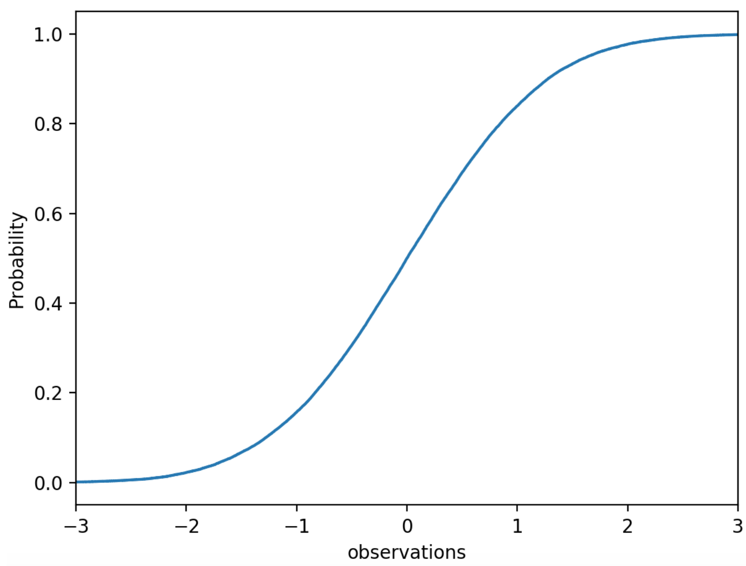

Figure 4: A.12a

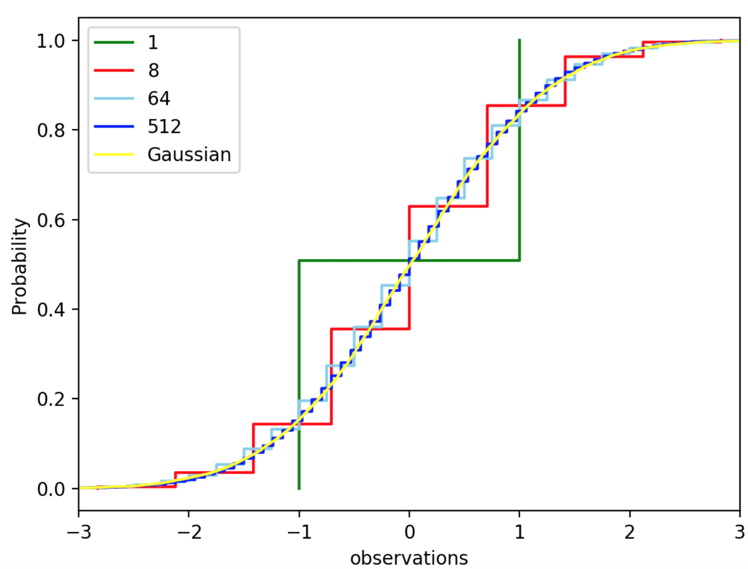

Figure 5: A.12b

[3]:
```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
%matplotlib notebook

# Create the vectors X and Y
x1 = np.array(range(-10,10))
x2 = x1/2 -1
# Create the plot
plt.plot(x1,x2,label='$x_1+2x_2+2=0$')
# Add a title
plt.title('A.9 a')
# Add X and y Label
plt.xlabel('$x_1$ axis')
plt.ylabel('$x_2$ axis')
# Add a grid
plt.grid(alpha=.4,linestyle='--')
# Add a Legend
plt.legend()
# Show the plot
plt.show()
```

[8]:
```
x1x1 = np.arange(-10,10,1)
x2x2 = np.arange(-10,10,1)
X, Y = np.meshgrid(x1x1, x2x2)
Z = -X-Y
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X,Y,Z,alpha=0.4)
ax.set_title('A.9 b')
ax.set_xlabel("$x_1$",fontsize=10)
ax.set_ylabel("$x_2$",fontsize=10)
ax.set_zlabel("$x_3$",fontsize=10)
#ax.legend(loc='x_1+x_2+x_3=1')
plt.show()
```

[4]:
```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
%matplotlib notebook
n=40000
Z=np.random.randn(n)
plt.step(sorted(Z), np.arange(1,n+1)/float(n))
plt.xlim(-3, 3)
# Add X and y Label
plt.xlabel('observations')
plt.ylabel('Probability')
plt.show()
```

[26]:
```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
%matplotlib notebook
n=10000
x_axix = np.arange(1,n+1)/float(n)
Z1 = np.sum(np.sign(np.random.randn(n, 1))*np.sqrt(1./1), axis=1)
Z2 = np.sum(np.sign(np.random.randn(n, 8))*np.sqrt(1./8), axis=1)
Z3 = np.sum(np.sign(np.random.randn(n, 64))*np.sqrt(1./64), axis=1)
Z4 = np.sum(np.sign(np.random.randn(n, 512))*np.sqrt(1./512), axis=1)
Z5 = np.random.randn(n)
plt.plot(sorted(Z1), x_axix,  color='green', label='1')
plt.plot(sorted(Z2), x_axix,  color='red', label='8')
plt.plot(sorted(Z3), x_axix, color='skyblue', label='64')
plt.plot(sorted(Z4), x_axix,  color='blue', label='512')
plt.plot(sorted(Z5), x_axix,  color='yellow', label='Gaussian')
plt.legend() #
plt.xlim(-3, 3)
plt.xlabel('observations')
plt.ylabel('Probability')
plt.show()
```