

# CSE546 HW0 A

Bobby Deng — 1663039 — dengy7

March 2020

## 1 A.1

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

$$P(B) = P(A) * P(B|A) + P(A^-) * P(B|A^-)$$

$$P(A^-) = 1 - P(A)$$

$A$ : You have the disease.  $A^-$ : You don't have the disease.

$B$ : Test is positive.  $B^-$ : Test is negative.

$$P(A) = 0.0001, P(A^-) = 0.9999$$

$$P(B) = P(A) * P(B|A) + P(A^-) * P(B|A^-)$$

$$P(B) = 0.0001 * 0.99 + 0.9999 * 0.01$$

$$P(B) = 0.010098$$

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

$$P(A|B) = \frac{0.0001 * 0.99}{0.010098}$$

$$P(A|B) = 0.0098$$

## 2 A.2

### 2.1 a.

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

$$Cov(X, Y) = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$$

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

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

1. **When**  $E[Y|X = x] = x$ , **then**  $E[X] = E[Y]$ , **so we can change all**  $E[Y]$  **to**  $E[X]$ .

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

$$Cov(X, Y) = E[XX] - E[X]E[X + X] + E[X]^2$$

$$Cov(X, Y) = E[XX - XE[X] - XE[X] + E[X]^2]$$

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

$$Cov(X, Y) = E[(X - E[X])^2]$$

## 2.2 b.

When X,Y are independent:

$$Cov(X, Y) = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$$

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

1. We know that  $E[XY] = E[X]E[Y]$ , Then

$$Cov(X, Y) = 0$$

## 3 A.3

### 3.1 a.

Since  $Z = X + Y$ , the probability function for Z should be joint probability of X and Y.

$$h(z) = P(Z); Z = X + Y$$

$Y = 0, X = Z; Y = 1, X = Z-1; Y = 2, X = Z-2; \dots; Y = Z-1, X = 1; Y = 0, X = Z$

Since X,Y are independent, then

$$h(z) = \sum_0^Z P(X = i, Y_k = Z - i)$$

$$h(z) = \sum_0^Z P(X = i) * P(Y = Z - i)$$

$$h(z) = \sum_0^Z f(x) * g(y)$$

Above is showing what happened for discrete variable. It is the same story for continuous variable, however there are infinite many of that variable X,Y combination. So, we need to use integral to represents the infinite many variables and the area under it which representing the probability of a single Z.

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

We know that  $Z = X + Y$ , so we can replace  $Y = Z - X$ , then it becomes:

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

### 3.2 b.

First of all, Z can not be less than 0 or bigger than 2 since X,Y are on [0,1].

Then we look at Z in [1,2] and Z in [0,1] separately.

$f(x) * g(z - x)$  Could either equal to 0 or 1.

When X and Z-x are both on the designated interval,  $f(x) * g(z - x) = 1$ , 0 otherwise.

Now we only look at situations where it is 1.

When  $1 \geq x \geq 0$  or  $1 \geq Z - X \geq 0$

Then we get two intervals:  $Z \geq X \geq 0$  and  $1 \geq X \geq Z - 1$

We Get:

$$h(z) = \int_0^z 1dx = z$$

$$h(z) = \int_{z-1}^1 1dx = 2 - z$$

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

## 4 A.4

We want  $E[Y] = 0$  and  $var(Y) = \sigma^2 = 1$

$$E[Y] = aE[X] + b = 0$$

$$0 = a\mu + b$$

$$b = -a\mu$$

$$var(Y) = var(aX + b) = 1$$

$$var(y) = a^2 var(X) + var(b) = 1$$

$$\text{var}(y) = a^2 \sigma^2 + 0 = 1$$

$$\text{var}(y) = a^2 \sigma^2 = 1$$

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

$$a = 1/\pm \sigma$$

So,

$$b = \mu/\pm \sigma$$

## 5 A.5

$$E[\hat{\mu}_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

$$\text{var}(\hat{\mu}_n) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$$

Now we know  $E[\hat{\mu}_n] = \mu$  and  $\text{var}(\hat{\mu}_n) = \frac{\sigma^2}{n}$ . We can use the above calculation for the equations below.

$$E[Z] = E[\sqrt{n}(\hat{\mu}_n - \mu)] = E[\sqrt{n}](E[\hat{\mu}_n] - E[\mu]) = \sqrt{n}(\mu - \mu) = 0$$

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

## 6 A.6

### 6.1 a.

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}$$

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

Lets work out what  $E[1\{X_i \leq x\}]$  is:

$$\begin{aligned} E[1\{X_i \leq x\}] &= \int_{-\infty}^{\infty} 1\{X_i \leq x\} * f(x) dx \\ &= \int_{-\infty}^{\infty} 1 * f(x) dx; \quad \text{when } X_i \leq x \\ &= F(x); \quad \text{when } X_i \leq x \end{aligned}$$

Now:

$$E[\hat{F}_n(x)] = \frac{1}{n} \sum_{i=1}^n E[1\{X_i \leq x\}] = \frac{1}{n} \sum_{i=1}^n F(x) = F(x)$$

## 6.2 b.

$$\begin{aligned} \text{var}(\hat{F}_n(x)) &= E[(\hat{F}_n(x) - F(x))^2] = E[(\hat{F}_n(x) - F(x))(\hat{F}_n(x) - F(x))] \\ &= E[\hat{F}_n(x)^2 - 2\hat{F}_n(x)F(x) + F(x)^2] \end{aligned}$$

$$\text{var}(\hat{F}_n(x)) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(1\{X_i \leq x\})$$

Lets work out what  $\text{var}(1\{X_i \leq x\})$  (Let A to represent it) is:

$$\text{var}(A) = E[(A - E[A])^2]$$

$$\text{var}(A) = E[A^2 - 2AE[A] + (E[A])^2]$$

Since A is 1, so  $A^2 = A$  :

$$\text{var}(A) = E[A] - 2E[A]E[A] + (E[A])^2]$$

$$\text{var}(A) = E[A] - E[A]E[A]$$

$$\text{var}(A) = F(x) - F(x)^2$$

$$\text{var}(A) = F(x)(1 - F(x))$$

$$\text{var}(\hat{F}_n(x)) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(1\{X_i \leq x\}) = \frac{1}{n^2} \sum_{i=1}^n F(x)(1 - F(x))$$

$$\text{var}(\hat{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}$$

### 6.3 c.

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

Prove if  $\frac{F(x)(1-F(x))}{n} \leq \frac{1}{4n}$ . Which is same as prove:  $F(x)(1 - F(x)) < \frac{1}{4}$ .

$$F(x)(1 - F(x)) = \frac{1}{4} - \frac{1}{4} + F(x)^2 - F(x)$$

$$F(x)(1 - F(x)) = \frac{1}{4} - \left(\frac{1}{2} - F(x)\right)^2 \leq \frac{1}{4}$$

Which is same as:

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

## 7 A.7

### 7.1 a.

Rank for A is: 2.

Rank for B is: 2.

### 7.2 b.

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

So the minimal basis for A:  $\begin{bmatrix} 1 \\ 0 \\ 1/2 \end{bmatrix}$  and  $\begin{bmatrix} 0 \\ 1 \\ 1/2 \end{bmatrix}$

$$B^T = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$$

So the minimal basis for B:  $\begin{bmatrix} 1 \\ 0 \\ 1/2 \end{bmatrix}$  and  $\begin{bmatrix} 0 \\ 1 \\ 1/2 \end{bmatrix}$

## 8 A.8

### 8.1 a.

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

**8.2 b.**

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

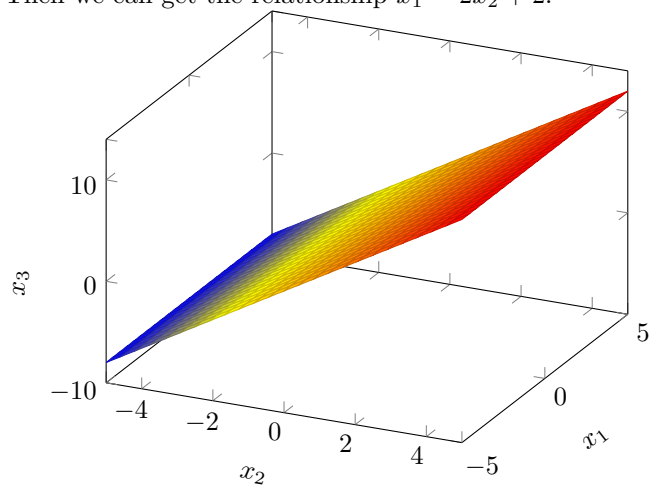
Solution to this augmented matrix is:  $\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$

**9 A.9**

**9.1 a.**

Based on the relationship, we can get this:  $\begin{bmatrix} -1 & 2 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2 = 0$

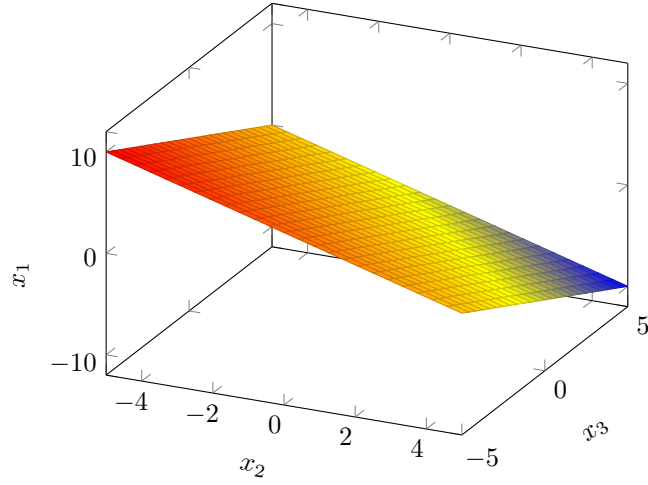
Then we can get the relationship  $x_1 = 2x_2 + 2$ .



**9.2 b.**

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + 0 = 0$$

Then we can get the relationship  $x_1 = -x_2 - x_3$ .



### 9.3 c.

When  $\tilde{x}_0$  is the minimizer,  $x = \tilde{x}_0$

$$\min_x \|x_0 - x\|^2 = \|x_0 - \tilde{x}_0\|^2 = \left| \frac{(w^T x_0 - w^T \tilde{x}_0)}{w^T w} \right|^2 = \left( \frac{w^T x_0 - w^T \tilde{x}_0}{w^T w} \right)^2$$

Since  $w^T x + b = 0$ , then  $w^T \tilde{x}_0 + b = 0$ ,  
Then  $b = -w^T \tilde{x}_0$ .

$$\left( \frac{w^T x_0 - w^T \tilde{x}_0}{w^T w} \right)^2 = \left( \frac{w^T x_0 + b}{w^T w} \right)^2$$

So the square distance is  $\left( \frac{w^T x_0 + b}{w^T w} \right)^2$ .

## 10 A.10

### 10.1 a.

$$\begin{aligned} x^T A x &= [x_1 \quad \dots \quad x_n] \begin{bmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= [\sum_{i=1}^n x_i A_{i,1} \quad \dots \quad \sum_{i=1}^n x_i A_{i,n}] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{j=1}^n (x_j (\sum_{i=1}^n x_i A_{i,j})) \end{aligned}$$

For  $y^T B x$ , it is the similar story,

$$y^T B x = \sum_{j=1}^n (x_j (\sum_{i=1}^n y_i B_{i,j}))$$



So,

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

**10.2 b.**

$$\nabla_z f(x, y) = \left[ \frac{\partial f(x, y)}{\partial z_1} \quad \dots \quad \frac{\partial f(x, y)}{\partial z_n} \right]^T$$

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

Let's look at each  $\frac{\partial f(x, y)}{\partial x_k}$  separately, and use k here in order to differentiate the i from matrix.

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

$$\begin{aligned}
\frac{\partial}{\partial x_k} &= \frac{\partial f(x, y)}{\partial x_k} \left( \sum_{j=1}^n (x_j \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j})) + c \right) \\
&= \sum_{j=1}^n \frac{\partial}{\partial x_k} [(x_j \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j})) + c] \\
&= \sum_{j=1}^n \frac{\partial}{\partial x_k} (x_j \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j})) + 0 \\
&= \sum_{j=1}^n \frac{\partial}{\partial x_k} (x_j \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j})) \\
&= \sum_{j=1}^n \left[ \frac{\partial}{\partial x_k} (x_j) \left( \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) + \frac{\partial}{\partial x_k} \left( \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) (x_j) \right] \\
&= \sum_{j=1}^n \frac{\partial}{\partial x_k} (x_j) \left( \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) + \sum_{j=1}^n \frac{\partial}{\partial x_k} \left( \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) (x_j) \\
&= [0 + 0 + \frac{\partial}{\partial x_k} (x_k) \left( \sum_{i=1}^n (x_i A_{i,k} + y_i B_{i,k}) \right) + \dots + 0 + 0] + [\sum_{j=1}^n (x_j \sum_{i=1}^n \frac{\partial}{\partial x_k} (x_i A_{i,j}) + 0)] \\
&= \sum_{i=1}^n (x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^n x_j (0 + 0 + \frac{\partial}{\partial x_k} (x_k A_{k,j}) + 0 + \dots + 0 + 0) \\
&= \sum_{i=1}^n (x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^n x_j \left( \frac{\partial}{\partial x_k} (x_k A_{k,j}) \right) \\
&= \sum_{i=1}^n (x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^n x_k A_{k,j} - \text{They are all from } 1 - n \\
&= \sum_{i=1}^n (x_i A_{i,k} + y_i B_{i,k} + x_k A_{k,i})
\end{aligned} \tag{1}$$

So,

$$\begin{aligned}
\nabla_x f(x, y) &= \left[ \frac{\partial f(x, y)}{\partial x_1} \quad \dots \quad \frac{\partial f(x, y)}{\partial x_n} \right]^T \\
&= \left[ \sum_{i=1}^n (x_i A_{i,1} + y_i B_{i,1} + x_1 A_{1,i}) \quad \dots \quad \sum_{i=1}^n (x_i A_{i,n} + y_i B_{i,n} + x_n A_{n,i}) \right]^T
\end{aligned} \tag{2}$$

### 10.3 c.

$$\nabla_y f(x, y) = \left[ \frac{\partial f(x, y)}{\partial y_1} \quad \dots \quad \frac{\partial f(x, y)}{\partial y_n} \right]^T$$

Now, do the same thing as x.

$$\begin{aligned}
 \frac{\partial f(x, y)}{\partial y_k} &= \frac{\partial}{\partial y_k} \left( \sum_{j=1}^n x_j \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) + c \\
 &= \sum_{j=1}^n x_j \left( \frac{\partial}{\partial y_k} \left( \sum_{i=1}^n (x_i A_{i,j} + y_i B_{i,j}) \right) \right) \\
 &= \sum_{j=1}^n x_j (0 + 0 + [0 + \frac{\partial}{\partial y_k} (y_k B_{k,j})] + 0 + \dots + 0) \\
 &= \sum_{j=1}^n x_j B_{k,j}
 \end{aligned} \tag{3}$$

So we can get,

$$\begin{aligned}
 \nabla_y f(x, y) &= \left[ \frac{\partial f(x, y)}{\partial y_1} \quad \dots \quad \frac{\partial f(x, y)}{\partial y_n} \right]^T \\
 &= \left[ \sum_{j=1}^n x_j B_{1,j} \quad \dots \quad \sum_{j=1}^n x_j B_{n,j} \right]^T
 \end{aligned} \tag{4}$$

## 11 A.11

### 11.1 a.

```

In [8]: # A.11
A = np.matrix([
    [0, 2, 4],
    [2, 4, 2],
    [3, 3, 1]
])
b = np.transpose([-2, -2, -4])
c = np.transpose([1, 1, 1])

# a.
np.linalg.inv(A)

Out[8]: matrix([[ 0.125, -0.625,  0.75 ],
                [-0.25 ,  0.75 , -0.5 ],
                [ 0.375, -0.375,  0.25 ]])

```

## 11.2 b.

```
In [10]: #b.  
np.linalg.inv(A).dot(b)
```

```
Out[10]: matrix([[ -2.,   1.,  -1.]])
```

```
In [11]: A.dot(c)
```

```
Out[11]: matrix([[6, 8, 7]])
```

## 12 A.12

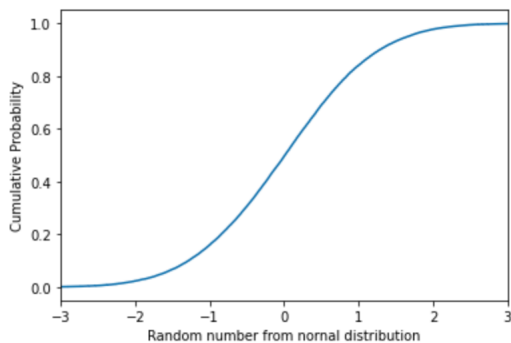
### 12.1 a.

According to previous answer:  $\text{var}(\hat{F}_n(x)) \leq \frac{1}{4n}$ , so:

$$\begin{aligned}\sqrt{\text{var}(\hat{F}_n(x))} &\leq 0.0025 \\ \text{var}(\hat{F}_n(x)) &\leq (0.0025)^2 \\ \text{var}(\hat{F}_n(x)) &\leq \frac{1}{4n} \\ (0.0025)^2 &= \frac{1}{4n} \\ n &= 4000\end{aligned}\tag{5}$$

```
In [5]: # a.  
n = 4000  
z = np.random.randn(n)  
plt.step(sorted(z), np.arange(1, n+1)/float(n))  
plt.xlabel('Random number from normal distribution')  
plt.ylabel('Cumulative Probability')  
plt.xlim(-3,3)
```

```
Out[5]: (-3.0, 3.0)
```



## 12.2 b.

```
In [18]: # b.  
ks = [1,8,64,512]  
  
for k in ks:  
    Z = np.sum(np.sign(np.random.randn(n,k))*np.sqrt(1./k), axis=1)  
    plt.step(sorted(Z), np.arange(1, n+1)/float(n), label=k)  
    plt.xlabel('Random number from normal distribution')  
    plt.ylabel('Cumulative Probability')  
    plt.xlim(-3,3)  
    plt.legend()
```

