CSE546 HW0

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.
$P(A) = 0.0.0001$, $P(A^-) = 0.9999$

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

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

$$P(B) = 0.010098$$

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

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

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

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.

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

$$Y = 0, X = Z; Y = 1, X = Z-1; Y = 2, X = Z-2; ....; Y = Z-1, X = 1; Y = 0, X = Z$$

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 $var(Y) = \sigma^2 = 1$

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

$$0 = a\mu + b$$

$$b = -a\mu$$

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

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

$$var(y) = a^2\sigma^2 + 0 = 1$$
$$var(y) = a^2\sigma^2 = 1$$
$$a^2 = 1/\sigma^2$$
$$a = 1/\pm\sigma$$

So,

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

5 A.5

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

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

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

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

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

6 A.6

6.1 a.

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

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

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

$$E[1\{X_i \leq x\}] = \int_{-\infty}^{\infty} 1\{X_i \leq x\} * f(x)dx$$

$$= \int_{-\infty}^{\infty} 1 * f(x)dx; \quad when X_i \leq x$$

$$= F(x); \quad when X_i \leq x$$

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.

$$var(\hat{F_n}(x)) = E[(\hat{F_n}(x) - F(x))^2] = E[(\hat{F_n}(x) - F(x))(\hat{F_n}(x) - F(x))]$$
$$= E[\hat{F_n}(x)^2 - 2\hat{F_n}(x)F(x) + F(x)^2]$$

$$var(\hat{F_n}(x)) = var(\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}) = \frac{1}{n^2}\sum_{i=1}^{n} var(1\{X_i \leq x\})$$

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

$$var(A) = E[(A - E[A])^2]$$

$$var(A) = E[A^2 - 2AE[A] + (E[A])^2]$$

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

$$var(A) = E[A] - 2E[A]E[A] + (E[A])^2]$$

$$var(A) = E[A] - E[A]E[A]]$$
$$var(A) = F(x) - F(x)^2$$
$$var(A) = F(x)(1 - F(x))$$

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

6.3 c.

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

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

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

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

Which is same as:

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

B.1

PDF is:

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

CDF is:

$$F_x(x) = \int_{-\infty}^{x} f(x)dx = x$$

Let me calculate the PDF of: $Y = Max(X_1, X_2, ..., X_n)$, and here we need to use joint probability to compute.

$$P(Max(X_1, X_2, ..., X_n) = P(X_1 \leq X, X_2 \leq X, ..., X_{n-1} \leq X)$$

So we have:

$$f_Y(x) = n[F_x(x)]^{n-1} f(x) = n(x)^{n-1} * 1 = n(x)^{n-1}$$

$$= \begin{cases} nx^{n-1}, & \text{for } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Now we need $E[Y]$, we use mean equation for continuous random variable.

$$E[Y] = \int_0^1 x f_M(x)dx = \int_0^1 xnx^{n-1}dx = n\int_0^1 x^n dx = \frac{n}{n+1}$$

7 A.7

7.1 a.

Rank for A is: 2.
Rank for B is: 2.

7.2 b.

Minimal size of basis for A: 2.
Minimal size of basis for B: 2.

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.

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

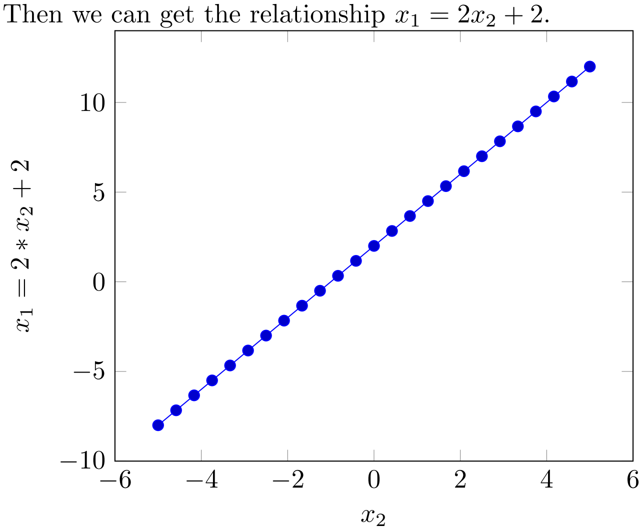
Solution to this augmented matrix is: $\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$

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

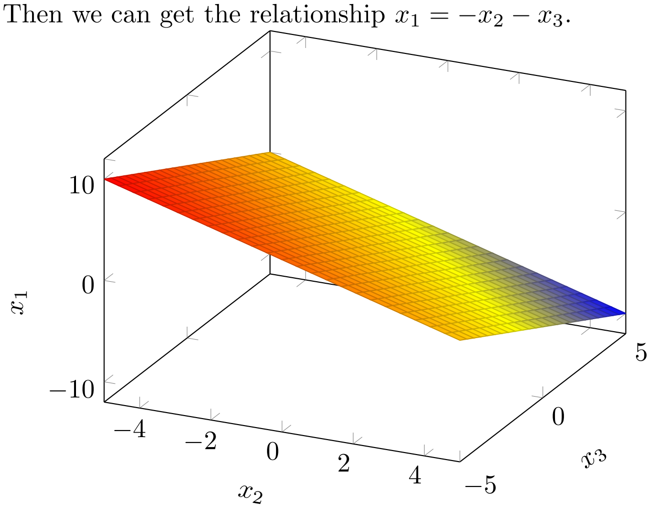
Then we can get the relationship $x_1 = 2x_2 + 2$.



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, $x = \tilde{x}_0$

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

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

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

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

10 A.10

10.1 a.

$$x^T A x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i A_{i,1} & \dots & \sum_{i=1}^{n} x_i A_{i,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{j=1}^{n} (x_j (\sum_{i=1}^{n} x_i A_{i,j}))$$

For $y^T Bx$, it is the similar story,

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

So,

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

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

10.2 b.

$$\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial z_1} & \dots & \frac{\partial f(x,y)}{\partial z_n} \end{bmatrix}^T$$

$$\nabla_x f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x_1} & \dots & \frac{\partial f(x,y)}{\partial x_n} \end{bmatrix}^T$$

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

$$
\begin{aligned}
\frac{\partial}{\partial x_k} &= \frac{\partial f(x,y)}{\partial x_k}(\sum_{j=1}^{n}(x_j\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))+c) \\
&= \sum_{j=1}^{n}\frac{\partial}{\partial x_k}[(x_j\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))+c] \\
&= \sum_{j=1}^{n}\frac{\partial}{\partial x_k}(x_j\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))+0 \\
&= \sum_{j=1}^{n}\frac{\partial}{\partial x_k}(x_j\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j})) \\
&= \sum_{j=1}^{n}[\frac{\partial}{\partial x_k}(x_j)(\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))+\frac{\partial}{\partial x_k}(\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))(x_j)] \\
&= \sum_{j=1}^{n}\frac{\partial}{\partial x_k}(x_j)(\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))+\sum_{j=1}^{n}\frac{\partial}{\partial x_k}(\sum_{i=1}^{n}(x_iA_{i,j}+y_iB_{i,j}))(x_j) \\
&= [0+0+\frac{\partial}{\partial x_k}(x_k)(\sum_{i=1}^{n}(x_iA_{i,k}+y_iB_{i,k}))+\cdots+0+0]+[\sum_{j=1}^{n}(x_j\sum_{i=1}^{n}\frac{\partial}{\partial x_k}(x_iA_{i,j})+0)] \\
&= \sum_{i=1}^{n}(x_iA_{i,k}+y_iB_{i,k})+\sum_{j=1}^{n}x_j(0+0+\frac{\partial}{\partial x_k}(x_kA_{k,j})+0+\cdots+0+0) \\
&= \sum_{i=1}^{n}(x_iA_{i,k}+y_iB_{i,k})+\sum_{j=1}^{n}x_j(\frac{\partial}{\partial x_k}(x_kA_{k,j})) \\
&= \sum_{i=1}^{n}(x_iA_{i,k}+y_iB_{i,k})+\sum_{j=1}^{n}x_kA_{k,j}-Theyareallfrom1-n \\
&= \sum_{i=1}^{n}(x_iA_{i,k}+y_iB_{i,k}+x_kA_{k,i})
\end{aligned} \tag{1}
$$

So,

$$
\begin{aligned}
\nabla_x f(x,y) &= \begin{bmatrix} \frac{\partial f(x,y)}{\partial x_1} & \ldots & \frac{\partial f(x,y)}{\partial x_n} \end{bmatrix}^T \\
&= \begin{bmatrix} \sum_{i=1}^{n}(x_iA_{i,1}+y_iB_{i,1}+x_1A_{1,i}) & \ldots & \sum_{i=1}^{n}(x_iA_{i,n}+y_iB_{i,n}+x_kA_{n,i}) \end{bmatrix}^T
\end{aligned} \tag{2}
$$

10.3 c.

$$
\nabla_y f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial y_1} & \ldots & \frac{\partial f(x,y)}{\partial y_n} \end{bmatrix}^T
$$

Now, do the same thing as x.

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

So we can get,

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

B.2

Since, $A \in \mathbb{R}^{\mathrm{n \times m}}$, and $B \in \mathbb{R}^{\mathrm{m \times n}}$ So,

$$\begin{aligned}
tr(AB) &= \sum_{i=1}^{m}(AB)_{ii} \\
&= \sum_{i=1}^{m}\sum_{j=q}^{n} A_{i,j} B_{j,i} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m} B_{j,i} A_{i,j} \\
&= \sum_{j=1}^{n}(BA)_{j,j} \\
&= tr(BA)
\end{aligned} \tag{5}$$

B.3

a.

When we consider a matrix of d by d, so,
Max rank: d and min rank: 0;

b.

Because V is a b by n matrix, so,
Max rank: d, min rank: 0;

c.

Since A is D by d and v_i is d by 1, so Av_i is D by 1. So, we could know that $(Av_i)^T$ is 1 by D.

So $(Av_i)(Av_i)^T$ is D by D. Then,
Max rank: D, min rank: 0.

d.

Since V is d by n, so AV is D by n, so,
If $D > n$, then max rank is n, min is 0,
If $n > D$, then max rank is D, min is 0.

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

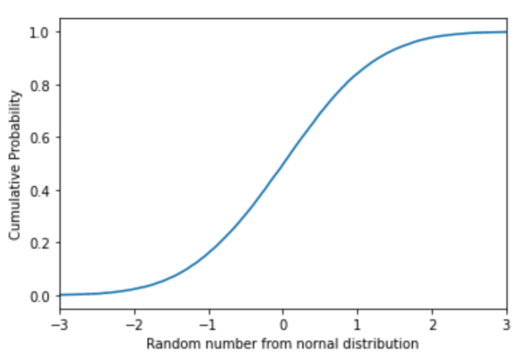
12.1 a.

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

$$\begin{aligned} \sqrt{var(\hat{F_n}(x))} \leq 0.0025 \\ var(\hat{F_n}(x)) \leq& (0.0025)^2 \\ var(\hat{F_n}(x)) \leq& \frac{1}{4n} \\ (0.0025)^2 =& \frac{1}{4n} \\ n =& 4000 \end{aligned} \tag{6}$$

```
In [5]: # a.
        n = 40000
        Z = np.random.randn(n)
        plt.step(sorted(Z), np.arange(1, n+1)/float(n))
        plt.xlabel('Random number from nornal 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 nornal distribution')
    plt.ylabel('Cumulative Probability')
    plt.xlim(-3,3)
    plt.legend()
```