EM 397 HW #4
Group 4 (Soovadeep Bakshi, William Ruys, Wenbo Zhang, Akshay Kumar Varanasi)

Ex. 12.1:

Ex. 12.1 (a)

By Markov's inequality:

$$\mathbb{P}[m \geq t] \leq \frac{\mathbb{E}[m]}{t}, \forall t > 0, \quad m = \text{non-negative r.v.}$$

Let m have zero mean and variance σ^2

To prove Chebyshev's inequality, we apply the Markov inequality to the r.v. $Y = m^2$,

We know that

$$\mathbb{P}[|m| \geq t] = \mathbb{P}[m^2 \geq t^2] \overset{\text{by Markov}}{\leq} \frac{\mathbb{E}[m^2]}{t^2} = \frac{\sigma^2}{t^2}, \forall t > 0$$

$$\Rightarrow \mathbb{P}[|m| \geq t] \leq \frac{\sigma^2}{t^2} \quad \forall t > 0$$

Ex 12.1 (b)

To prove the Chernoff bound, we use the fact that the exponential is monotonic

$$\Rightarrow \mathbb{P}[m \geq t] = \mathbb{P}[e^{\lambda m} \geq e^{\lambda t}] \leq \frac{\mathbb{E}[e^{\lambda m}]}{e^{\lambda t}} \quad \forall \lambda > 0, \ t > 0 \quad (*)$$

Thus, since $(*)$ is valid for every $\lambda > 0$, we can optimize over λ to get

$$\mathbb{P}[m \geq t] \leq \min_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda m}]}{e^{\lambda t}}, \quad \forall t > 0$$

Ex 12.2:

Ex. 12.2

1) Given $m \sim N(\bar{m}, \sigma^2)$

By definition,

$$M(t) = \mathbb{E}[e^{tm}] = \int_{-\infty}^{\infty} e^{tm}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(m-\bar{m})^2}{2\sigma^2}\right)\right]dm \qquad \forall t \in \mathbb{R}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2}\left(m^2 - 2m\bar{m} + \bar{m}^2 - 2\sigma^2 t m\right)\right]dm$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2}\left(\left(m - (\bar{m}+\sigma^2 t)\right)^2 - (\bar{m}+\sigma^2 t)^2 + \bar{m}^2\right)\right]dm$$

$$\left[\begin{array}{l} \text{using the fact: } m^2 - 2cm + d = (m-c)^2 - c^2 + d \\ c = (\bar{m}+\sigma^2 t),\ d = \bar{m}^2 \end{array}\right]$$

$$= \exp\left[\frac{-1}{2\sigma^2}\left(-(\bar{m}+\sigma^2 t)^2 + \bar{m}^2\right)\right] \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(m-(\bar{m}+\sigma^2 t))^2}{2\sigma^2}\right)dm}_{m' \sim N(\bar{m}+\sigma^2 t,\ \sigma^2) \ \Rightarrow \int_{-\infty}^{\infty} \cdots\, dm = 1}$$

$$= \exp\left[\frac{\bar{m}^2 + 2\bar{m}\sigma^2 t + \sigma^4 t^2 - \bar{m}^2}{2\sigma^2}\right] = \exp\left[\bar{m}t + \frac{\sigma^2 t^2}{2}\right]$$

$\forall t \in \mathbb{R}$

~~2) $\min_{\lambda>0} \frac{\mathbb{E}[e^{\lambda m}]}{\mathbb{E}e^{\lambda t}} = e^{\lambda\bar{m} + \frac{\sigma^2\lambda^2}{2} - \lambda t}$~~

~~This is equivalent to the minimization problem:~~

2) Given, $m \sim N(\bar{m}, \sigma^2) \Rightarrow m - \bar{m} \sim N(0, \sigma^2)$

Therefore, applying the Chernoff bound and the MGF of a normal r.v,

$$\mathbb{P}[m - \bar{m} \geq t] \leq \min_{\lambda>0} \frac{\mathbb{E}[e^{\lambda(m-\bar{m})}]}{e^{\lambda t}} = \min_{\lambda>0} \exp\left(\frac{\sigma^2\lambda^2}{2} - \lambda t\right), \quad \forall t > 0$$

This is equivalent to the minimization problem.

$$\min_{\lambda>0} \frac{\sigma^2\lambda^2}{2} - \lambda t \qquad \text{(since exponential is monotonic)}$$

$\Rightarrow \lambda = \frac{t}{\sigma^2}$ is the minimizer (solved by setting $\frac{\partial F(\lambda)}{\partial\lambda} = 0$)

$$\Rightarrow \mathbb{P}[m - \bar{m} \geq t] \leq e^{-\frac{t^2}{\sigma^2} + \frac{t^2}{2\sigma^2}} = e^{-\frac{t^2}{2\sigma^2}}, \quad \forall t > 0$$

Ex 14.1:

Ex 14.1

Let m be a Rademacher r.v., i.e. $m = 1$ or -1 with probability $1/2$

To prove that the Rademacher distribution is sub-gaussian,

$$\mathbb{E}\left[e^{\lambda m}\right] = \frac{e^{\lambda} + e^{-\lambda}}{2} = \cosh(\lambda)$$

$$= \sum_{n \geq 0} \frac{\lambda^{2n}}{(2n)!} \leq \sum_{n \geq 0} \frac{\lambda^{2n}}{2^n n!} \quad \left(\text{using } 2^n n! \leq (2n)!\right)$$

$$\sum_{n \geq 0} \frac{\lambda^{2n}}{2^n n!} = \exp\left(\frac{\lambda^2}{2}\right)$$

$$\Rightarrow \mathbb{E}\left[e^{\lambda m}\right] \leq e^{\frac{\lambda^2}{2}}$$

Ex. 14.2:

Ex. 14.2

Using the convexity of the exponential function and the Jensen's ~~eto~~ inequality ($f(\mathbb{E}[z]) \leqslant \mathbb{E}[f(z)]$ ~~where~~ if f is convex), we can write the following:

$$\mathbb{E}_m\left[e^{\lambda m}\right] = \mathbb{E}_m\left[e^{\lambda(m-\mathbb{E}_m[m])}\right] = \mathbb{E}_m\left[e^{\lambda(m-\mathbb{E}_{m'}[m'])}\right] \underset{\text{(Jensen)}}{\leqslant} \mathbb{E}_m\left[\mathbb{E}_{m'}\left[e^{\lambda(m-m')}\right]\right]$$

where m' is an independent copy of m

$\Rightarrow m' \in [a,b]$, $\mathbb{E}_m[m] = \mathbb{E}_{m'}[m'] = 0$

Now, the difference $m-m'$ is symmetric about 0. Therefore, we can show that for the r.v. $\tilde{m} = m - m'$, $y = \{-1, 1\}$, $\tilde{m}y$ has the same distribution as $\tilde{m}$ ($\tilde{m}$ and y are independent)

Proof:

$$\mathbb{P}[\tilde{m}y \leqslant a] = \mathbb{P}[\tilde{m} \leqslant a \mid y=1].\mathbb{P}[y=1] + \mathbb{P}[\tilde{m} \leqslant a \mid y=-1].\mathbb{P}[y=-1]$$

$\parallel$ using symmetry and independence

$$\mathbb{P}[\tilde{m} \leqslant a].(\mathbb{P}[y=1] + \mathbb{P}[y=-1])$$

$$\parallel$$

$$\mathbb{P}[\tilde{m} \leqslant a]$$

$\Rightarrow$ $\tilde{m}y$ has the same distribution as $\tilde{m}$

$$\Rightarrow \mathbb{E}_{m,m'}\left[e^{\lambda(m-m')}\right] = \mathbb{E}_{m,m'}\left[\mathbb{E}_y\left[e^{\lambda y(m-m')}\right]\right]$$

where y is a Rademacher r.v. Thus, using (14.1),

$$\mathbb{E}_y\left[e^{\lambda y(m-m')} \mid m, m'\right] \leqslant \exp\left(\frac{\lambda^2(m-m')^2}{2}\right)$$

By assumption, we have $|m-m'| \leqslant (b-a) \Rightarrow (m-m')^2 \leqslant (b-a)^2$

$$\Rightarrow \mathbb{E}_{m,m'}\left[e^{\lambda(m-m')}\right] \leqslant \exp\left(\frac{\lambda^2(b-a)^2}{2}\right)$$

$$\Rightarrow \mathbb{E}_m\left[e^{\lambda m}\right] \leqslant \exp\left(\frac{\lambda^2(b-a)^2}{2}\right)$$

Ex 14.3:

Exercise 14.3.

Let m_i , $i=1,2,..N$ be independent bounded random variables.
$m_i \in [a_i, b_i]$ a.s.

Show $\mathbb{P}[\,|S_N - \mathbb{E}[S_N]| > \epsilon] \leq 2e^{\frac{-2N\epsilon^2}{\Sigma(b_i-a_i)^2}}$ $\forall \epsilon > 0$.

$$\mathbb{P}[S_N - \mathbb{E}[S_N] \geq \epsilon] \leq \min_{\lambda > 0} \frac{\mathbb{E}\left[e^{\lambda(S_N - \mathbb{E}[S_N])}\right]}{e^{\lambda t}}$$

$$e^{-\lambda t}\,\mathbb{E}\left[e^{\lambda(S_N - \mathbb{E}[S_N])}\right] = e^{-\lambda t}\,\mathbb{E}\left[e^{\lambda \frac{1}{N}(\Sigma m_i - \mathbb{E}[m_i])}\right]$$

(by independence) $= e^{-\lambda t} \prod_{i=1}^{N} \mathbb{E}\left[e^{\lambda/N\,(m_i - \mathbb{E}[m_i])}\right]$

By 14.2 $\mathbb{E}\left[e^{\lambda/N\,(m_i - E[m_i])}\right] \leq e^{\frac{\lambda^2/N^2\,(b_i - a_i)^2}{8}}$
as m_i bdd and $\mathbb{E}[m_i - E[m_i]] = 0$. using 14.2.

so all together:

$$\mathbb{P}[S_N - \mathbb{E}[S_N] \geq t] \leq \min_{\lambda > 0} e^{-\lambda t}\, e^{\frac{\lambda^2}{N^2}\sum_{j=1}^{N}(b_j - a_j)^2/8}$$

$$= \min_{\lambda > 0} e^{-\lambda t + \frac{\lambda^2}{8N^2}\Sigma(b_j - a_j)^2}$$

This is parabolic in $\lambda \Rightarrow$ minimum at axis of symmetry

$$\lambda_{min} = \frac{-b}{2a} = \frac{t}{2\left(\frac{\Sigma(b_j - a_j)^2}{8N^2}\right)} = \frac{4N^2 t}{\Sigma(b_j - a_j)^2}$$

$$e^{-\lambda_{min} t} = e^{\frac{-4Nt^2}{\Sigma(b_j - a_j)^2}}$$

$$e^{\frac{\lambda_{min}^2\,\Sigma(b_j - a_j)^2}{N^2 8}} = e^{\frac{2N^2 t^2}{\Sigma(b_j - a_j)^2}}$$

$$\Rightarrow \mathbb{P}[S_N - \mathbb{E}[S_N] \geq t] \leq e^{-2N^2 t^2/\Sigma(b_j - a_j)^2}$$

$\mathbb{P}[\ S_N - \mathbb{E}[S_N] \leq -t\ ] \leq e^{-2N^2t^2 / \Sigma(b_j - a_j)^2}$

by symmetry Take $m_i = -m_i \in [b_i, -a_i]$

$\Rightarrow \mathbb{P}[\ |S_N - \mathbb{E}[S_N]| \geq t\ ] \leq 2e^{-2N^2t^2 / \Sigma(b_j - a_j)^2}$.

- For $M_N = \Sigma m_i$ then

  $\mathbb{P}[\ |M_N - \mathbb{E}[M_N]| \geq t\ ] \leq 2e^{-2t^2 / \Sigma(b_j - a_j)^2}$.

- Toss a fair coin N times, what is probability of getting at least $3N/4$ heads. $\{$heads $= 1$, tails $= 0\}$

with bound.

$$\begin{cases} E[M_N] = N/2 \qquad t = N/4 \\ \mathbb{P}[\ M_N - N/2 \geq 1/4\ ] \leq e^{-\frac{2N^2}{4^2 N}} \leq e^{\frac{-N}{8}} \end{cases}$$

Ex 14.4:

Exercise 14.4

Def^{n} $s = 1/1-\varepsilon \qquad \varepsilon \in [0,1)$

$$\xi = \sqrt{s} \begin{cases} +1 & 1/2s & ① \\ 0 & \varepsilon = 1 - 1/s & ② \\ -1 & 1/2s & ③ \end{cases}$$

Show $\mathbb{E}[\xi] = 0$

$$\mathbb{E}[\xi] = \xi_1 P(1) + \xi_2 P(2) + \xi_3 P(3)$$

$$= \sqrt{s}\,\frac{1}{2s} + 0(1 - 1/s) - \sqrt{s}\,\frac{1}{2s} = 0 \quad \checkmark$$

$$\mathbb{E}[\xi^2] = s\left(\frac{1}{2s}\right) + 0 + s\left(\frac{1}{2s}\right) = 1 \quad \checkmark$$

Show ξ is subgaussian.

Idea: Since $E[\xi] = 0$ show $E[e^{\lambda\xi}] \le e^{\sigma^2\lambda^2/2}$.

$$E[e^{\lambda\xi}] = e^{\lambda\sqrt{s}}\left(\frac{1}{2s}\right) + e^{\lambda 0}\left(1 - 1/s\right) + e^{-\lambda\sqrt{s}}\left(1/2s\right)$$

$$= \frac{1}{2s}\left(e^{\lambda\sqrt{s}} + e^{-\lambda\sqrt{s}}\right) + \left(1 - 1/s\right)$$

$$= \frac{1}{s}\cosh(\lambda\sqrt{s}) + 1 - \frac{1}{s}$$

$$= \frac{1}{s}\sum_{n \ge 0}\frac{(\lambda\sqrt{s})^{2n}}{(2n)!} + 1 - \frac{1}{s}$$

$$= \frac{1}{s}\sum_{n \ge 1}\frac{\lambda^{2n}s^n}{(2n)!} + 1$$

$$\le \sum_{n \ge 1}\frac{\lambda^{2n}s^n}{(2n)!} + 1 \le \sum_{n \ge 1}\frac{\lambda^{2n}s^n}{2^n n!} + 1$$

$$= e^{\frac{\lambda^2 s}{2}} - 1 + 1 = e^{\lambda^2 s/2}$$

$$\Rightarrow \quad \sigma_i^2 = s.$$

Ex 14.5:

Exercise 14.5

Prove a concentration result like:

$$\mathbb{P}\left[\left| \tfrac{1}{N} \sum_{i=1}^{N} m_i \right| > t \right] \le 2 e^{-N t^2 / c^2} \quad \forall t \ge 0$$

when $m_i - \bar{m}_i$ are subgaussian w/ proxy σ_i^2

$$m_i - \bar{m}_i \text{ subgaussian} \Rightarrow \mathbb{E}\left[e^{\lambda (m_i - \bar{m}_i)} \right] \le e^{\lambda^2 \sigma_i^2 / 2}$$

Let's start with the general Hoeffding inequality.

$$\mathbb{P}\left[\left| \sum_{i=1}^{N} a_i (m_i - \bar{m}_i) \right| \ge t \right] \le \min_{\lambda > 0} \frac{\mathbb{E}\left[e^{\lambda \sum_{i=1}^{N} a_i (m_i - \bar{m}_i)} \right]}{e^{\lambda t}}$$

(Assume independence of $m_i - \bar{m}_i$)

$$= \min_{\lambda > 0} e^{-\lambda t} \prod_{i=1}^{N} \mathbb{E}\left[e^{\lambda a_i (m_i - \bar{m}_i)} \right]$$

$$\le \min_{\lambda > 0} e^{-\lambda t} \prod_{i=1}^{N} e^{\lambda^2 a_i^2 \sigma_i^2 / 2}$$

$$= \min_{\lambda > 0} e^{-\lambda t} e^{\sum_{i=1}^{N} \frac{\lambda^2 a_i^2 \sigma_i^2}{2}} = \min_{\lambda > 0} e^{-\lambda t + \frac{\lambda^2}{2} \sum_{i=1}^{N} a_i^2 \sigma_i^2}$$

$-\lambda t + \frac{\lambda^2}{2} \left(\sum_{i=1}^{N} a_i^2 \sigma_i^2 \right)$ is parabolic (w/ $a > 0$)

$$\Rightarrow \lambda_{min} = \frac{b}{a} \quad \Rightarrow \min = \frac{-b^2}{2a}$$

$$\Rightarrow \le \exp\left[\frac{-t^2}{\sum_{i=1}^{N} a_i^2 \sigma_i^2} \right]$$

So for $a_i = 1/N$

$$\mathbb{P}\left[\left| \frac{1}{N} \sum_{i=1}^{N} (m_i - \bar{m}_i) \right| > t \right] \le e^{-t^2 / \sum_{i=1}^{N} \left(\frac{1}{N^2}\right) \sigma_i^2}$$

$$= e^{-N^2 t^2 / \sum_{i=1}^{N} \sigma_i^2}$$

Ex 15.1:

We implemented the algorithm from [4]. (Bai and Golub's paper)
Algorithm 2 and Algorithm1 from the paper are given on the next page.
Note we use N=k^2 = 100 so we can easily compute the trace directly. A reasonable crossover point isn't until N>10,000s with the sampling parameter given in the book.

```
%Setup
k = 10;
N = k^2;
v = 0.2;
o = ones(1, k-1)*-v;
d = ones(1, k)*(1+4*v);
D = sparse(diag(d, 0)+diag(o, -1)+diag(o, 1));
C = sparse(diag(ones(1, k)*-v, 0));
A = blktridiag(D, C, C, k)';
f = @(x) x; %Just the identity of the eigenvalues bc we consider tr(A)
%Set this to f=@(x) 1./x and take a=1/b b=1/a to compute tr(A^-1)
%Estimate a and b from gershgorin circles
for i=1:size(A, 1)
      r(i) = sum(A(i, :)) - A(i, i);
      upper(i) = A(i, i)+r(i);
      lower(i) = A(i, i)-r(i);
end
a = min([upper, lower]);
b = max([upper, lower]);
if a<0
      a = 1e-4;
end
p = 0.5; %Probability it's outside the bounds
[Ip, Lp, Up] = Algorithm2(A, f, a, b, p);
```

We also sampled the Umax and Umin directly.

```
%Estimates from 333 million samples (took a lot of cores to do in reasonable time)
Umax = 212.8000;
Lmin = 145.6000;

%Estimates from 1000 samples
Umax = 194.4;
Lmin = 164.8;

%Lp, Up from 1000 samples, p = 0.9
Up = 180.4653;
Lp = 178.6995;

%Lp, Up from 1000 samples, p = 0.5
%Up = 180.2695
%Lp = 179.3849
```

```
%Algorithm to compute the bounds
function [U, L] = Algorithm1(u, A, f, a, b)
      x{2} = u./norm(u, 2);
      x{1} = 0;
      I{2, 1} = 0;
      I{2, 2} = 0;
      gamma(2) = 0;
      notconverged = 1;
      j = 3;
      while notconverged
      alpha(j) = x{j-1}'*A*x{j-1};
      r{j} = A*x{j-1} - alpha(j)*x{j-1} - gamma(j-1)*x{j-2};
      gamma(j) = norm(r{j}, 2);
      o = 2; %offset
      i = 1;
      T = diag(alpha(3:j), 0) + diag(gamma(3:j-1), 1) + diag(gamma(3:j-1), -1);
      for c=[a, b]
            Y = eye(j-o);
            delta = (T - c*eye(j-o, j-o))\(gamma(j)^2*Y(:, j-o));
            phi = delta(end);
            G = [T, gamma(j)*Y(:, j-o); gamma(j)*Y(:, j-o)', phi];
            [V, D] = eig(G);
            D = diag(D);
            V = V(1, :)';
            I{j, i} = sum((V.^2).*f(D));
            i = i+1;
      end
      if (j>10 || norm(I{j, 1} - I{j-1, 1}, 2)<1e-2)
            notconverged = false;
      end
      x{j} = r{j}./gamma(j);
      j = j+1;
      end
      U = norm(u, 2)^2*I{j-1, 1};
      L = norm(u, 2)^2*I{j-1, 2};
end
%Algorithm to compute the trace.
function [Ip, Lp, Up] = Algorithm2(A, f, a, b, p)
      N = size(A, 1);
      Lmin = 1e5;
      Umax = 0;
      m = 1000;
      for j=1:m
      z{j} = 2*(rand(N, 1)>0.5) - ones(N, 1);
      [L{j}, U{j}] = Algorithm1(z{j}, A, f, a, b);
      I{j} = 1/(2*j)*sum(cell2mat(L(:)) + cell2mat(U(:)));
      Lmin = min(Lmin, L{j});
      Umax = max(Umax, U{j});
      n = -0.5*j*(Umax - Lmin)^2*(log(1-p)/2);
      n = sqrt(n);
      Lp{j} = 1/j*sum(cell2mat(L(:))) - n/j;
      Up{j} = 1/j*sum(cell2mat(U(:))) + n/j;
      end
      Ip = I{m};
      Lp = Lp{m};
      Up = Up{m};
end
```

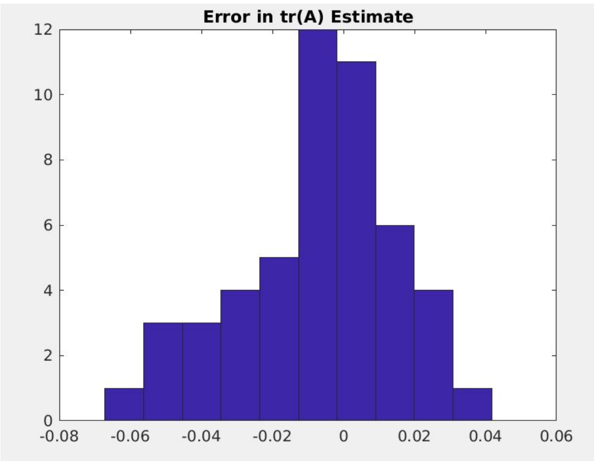
Now that we had the bounds we could compute the minimum number of samples.
As tr(A) = 180 (when A is 100x100) we computed the error directly.

```
n = ceil((Umax - Lmin)^2/(2*t^2)*log(2/(1-b)))
%Compute the estimator 50 times so we can compute experimental failure probability
for i = 1:50
      e = zeros(1, n);
      %Compute each estimator
      parfor j=2:n
      z = 2*(rand(N, 1)>0.5) - ones(N, 1);
      e(j) = z'*A*z;
      end
      I = sum(e)/n;
      err(i) = I - 180; %Subtract the true trace.
end

sum(abs(err)<t)/length(err) %Percent of estimates that meet the threshold
mean(err) %The average discrepancy
```

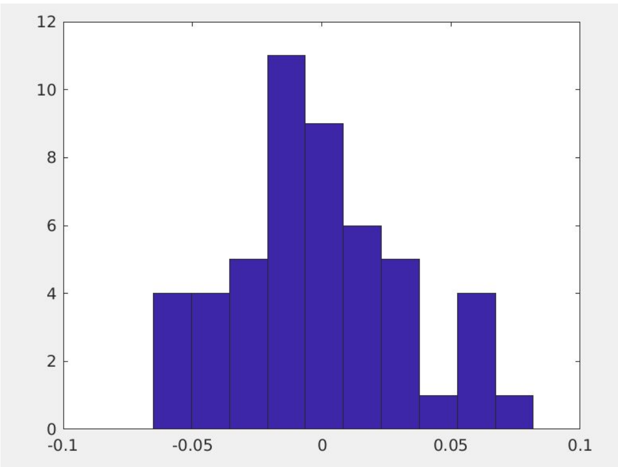
First, Algorithm 2 with p=0.5 reports:

```
Umax = 197.6000
Up = 180.5535
Ip = 180.1112
Lp = 179.6689
Lmin = 164.0000
```

Second we consider (t = 1e-1; b = 0.5)
Using `Umax = 194.4; Lmin = 164.8` **from a previous run we compute n=60731**
This gives an average underestimate of -0.0065 and distribution pictured below.



Even when we take it to the extreme (t=1e-1, beta=0.01) case. (n=30806) empirical success=1.



This means that the error bound for the minimum sample is quite loose and could be improved. In our experiments we actually saw average error on the order of t when we took:
`N = ceil((Umax - Lmin)^2/(2*t)*log(2/(1-b)))` **instead.**

Now for the inverse of A.
As tr(A^-1) = 58.3107 (when A is 100x100) we computed the error directly.

Algorithm 2 with p=0.5 reports:

```
Umax = 67.1931
Up =58.5245
Ip = 58.3472
Lp = 58.1699
Lmin = 52.9524
```

Ex 15.2:

15.2. $E[z_i]=0 \qquad Var[z_i] = \frac{\|x\|^2}{n}$

To show $E[\|z\|^2] = \|x\|^2$

A. $\|z\|^2 = \sum_{i=1}^{N} z_i^2$

$$E[\|z\|^2] = E\left[\sum_{i=1}^{N} z_i^2\right]$$

$$= \sum_{i=1}^{N} E[z_i^2] = \sum_{i=1}^{N} E[(z_i-0)^2]$$

$$\left[\because Var[z_i] = E[(z_i-0)^2] = \frac{\|x\|^2}{n}\right]$$

$$= \sum_{i=1}^{N} \frac{\|x\|^2}{n} = \cancel{n} \cdot \frac{\|x\|^2}{\cancel{n}}$$

$$E[\|z\|^2] = \|x\|^2$$

15-3

Ex 15.3:

15.3 Suppose $\xi \sim N(0, \sigma^2)$ and $t \leq 1/(2\sigma^2)$. Show that

$$\mathbb{E}\left[e^{t\xi^2}\right] = \frac{1}{\sqrt{1-2t\sigma^2}}$$ then deduce that

$$\mathbb{E}\left[e^{\lambda n z_i^2}\right] = \frac{1}{\sqrt{1-2\lambda\|x\|^2}} \quad \text{for } \lambda \leq 1/(2\|x\|^2)$$

A. $$\mathbb{E}\left[e^{t\xi^2}\right] = \int_{-\infty}^{\infty} e^{t\xi^2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\xi^2/2\sigma^2}\, d\xi = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\xi^2(2\sigma^2 t - 1)}{2\sigma^2}}\, d\xi$$

Let $\theta = \sqrt{1-2\sigma^2 t}\, \frac{\xi}{\sigma}$

$$d\theta = \frac{\sqrt{1-2\sigma^2 t}}{\sigma}\, d\xi$$

$$\mathbb{E}\left[e^{t\xi^2}\right] = \int_{-\infty}^{\infty} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-\theta^2/2} \frac{\cancel{\sigma}}{\sqrt{1-2\sigma^2 t}}\, d\theta$$

$$= \frac{1}{\sqrt{1-2\sigma^2 t}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\theta^2/2}\, d\theta}_{=1}$$

$$\mathbb{E}\left[e^{t\xi^2}\right] = \frac{1}{\sqrt{1-2\sigma^2 t}}$$

$z_i^2 \sim N(0, \frac{\|x\|^2}{n})$

$nz_i^2 \sim N(0, \|x\|^2)$

Substituting nz_i^2 for ξ_i^2 and $\|x\|^2$ for σ^2
$t \to \lambda$.

$$E\left[e^{\lambda n z_i^2}\right] = \frac{1}{\sqrt{1-2\lambda\|x\|^2}} \quad \text{for } \lambda \leq \frac{1}{2\|x\|^2}$$ //

Ex 15.4:

15.4 Show that

$$P\left[\|z\|^2 \le (1-\varepsilon)\|x\|^2\right] \le e^{n/2\left(-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\right)}$$

A $$P\left[-\lambda\|z\|^2 \ge (\varepsilon-1)\|x\|^2\right]$$

$$= \min e^{-n\lambda(\varepsilon-1)\|x\|^2}\, \Pi\, E\left[e^{-\lambda n z_1^2}\right]$$

$$= \min e^{n/2 f(\lambda)}$$

$$f(\lambda) = -2\lambda(\varepsilon-1)\|x\|^2 - \ln\left(1+2\lambda\|x\|^2\right) \qquad 0 \le \lambda \le \frac{1}{2\|x\|^2}$$

$$\frac{\partial f(\lambda)}{\partial \lambda} = -2(\varepsilon-1)\|x\|^2 - \frac{2\|x\|^2}{1+2\lambda\|x\|^2} = 0$$

$$\Rightarrow (1-\varepsilon) = \frac{1}{2\lambda^*\|x\|^2+1}$$

$$\Rightarrow 2\lambda^*\|x\|^2+1 = \frac{1}{1-\varepsilon}$$

$$\Rightarrow \lambda^* = \frac{1}{2\|x\|^2}\left[\frac{1}{1-\varepsilon}-1\right] = \frac{\varepsilon}{2\|x\|^2(1-\varepsilon)}$$

$$f(\lambda^*) = -2\frac{\varepsilon}{2\|x\|^2(1-\varepsilon)}(\varepsilon-1)\|x\|^2 - \ln\left(1+\frac{2\varepsilon}{2\|x\|^2(1-\varepsilon)}\|x\|^2\right)$$

$$= \varepsilon - \ln\left(\frac{1}{1+\varepsilon}\right) = \varepsilon + \ln(1-\varepsilon)$$

$$= \varepsilon - \varepsilon - \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} =$$

$$= -\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}$$

$$-\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} \leq -\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}$$

$$\therefore P\left[\|Z\|^2 \leq (1-\varepsilon)\|X\|^2\right] \leq e^{\frac{n}{2}\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)}$$

Using union bound

$$P\left[\|Z\|^2 \leq (1-\varepsilon)\|X\|^2 \text{ or } \|Z\|^2 \geq (1+\varepsilon)\|X\|^2\right] \leq 2e^{n/2\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)}$$

Ex 15.5:

15.5 Given $m \sim \chi_n^2$ then x can be represented as $m = \sum_{i=1}^{n} \xi_k^2$ where $\xi_k \sim N(0,1)$. Show that m concentrates around its mean with tail bound given by (15.2).

A We have $z = \lambda A x$ λ = scaling

let $x = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ $A = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \\ \vdots & & \end{bmatrix}$

$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$

For $\|z\|^2 \sim \chi_n^2$

$z_i = \lambda \sum_{i=1}^{n} a_{in} \sim N(0,1)$ since $a_{in} \sim N(0,1)$

z_i needs to be
$z_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{in}$ for it to be $N(0,1)$

$\Rightarrow P = \frac{1}{\sqrt{n}} A$ $x = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

since $\|z\|^2 \sim \chi^2_n$ we can apply 15.2

$$P\left[m \leq (1-\varepsilon)\bar{m} \text{ or } m \geq (1+\varepsilon)\bar{m} \right] \leq 2e^{\frac{n}{2}\left(-\frac{\varepsilon^2}{2}+\frac{\varepsilon^3}{3}\right)}$$

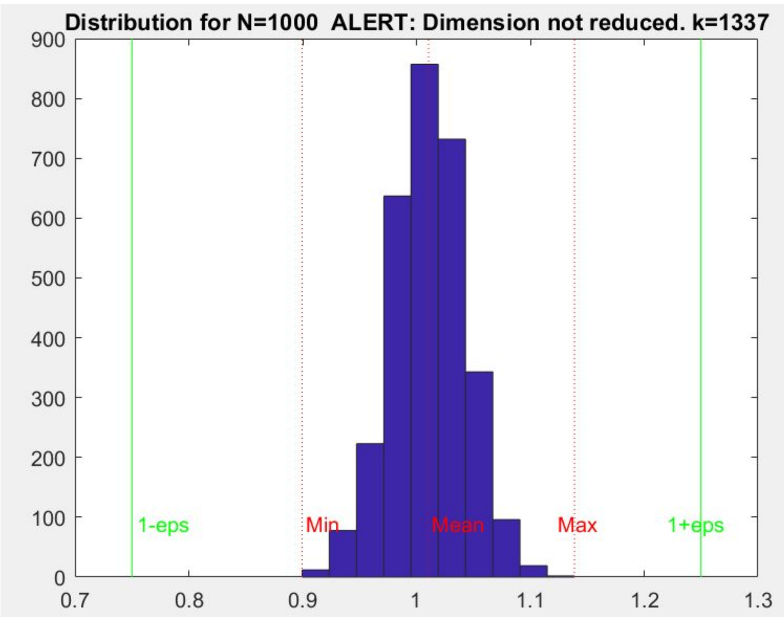
where $\bar{m}$ = mean

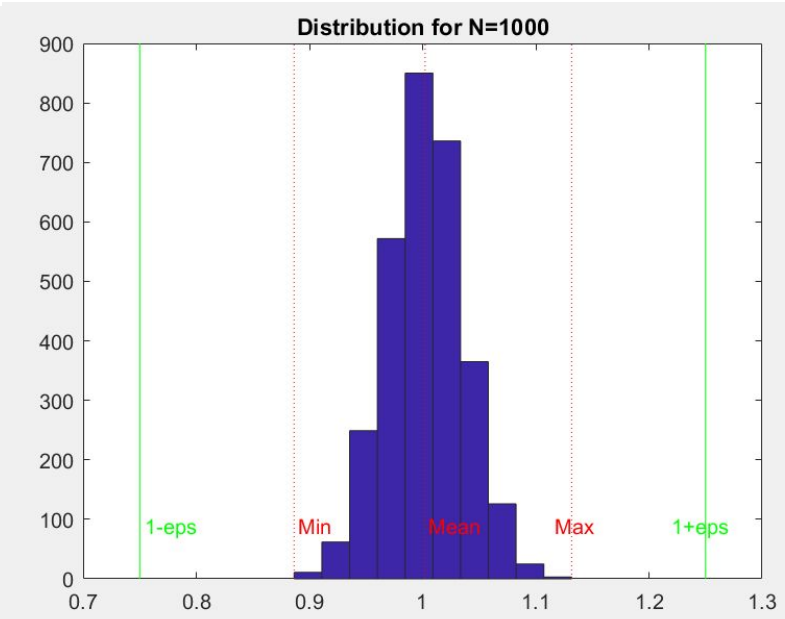
Ex 15.6:

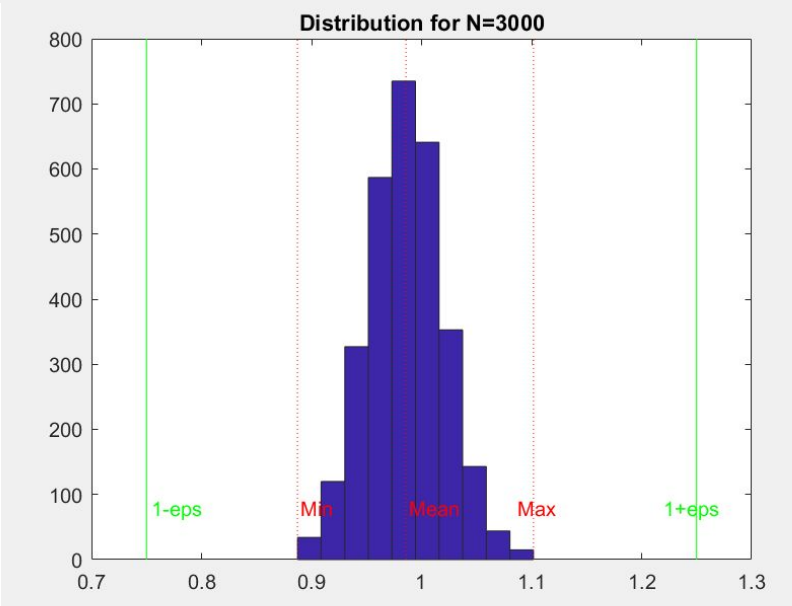
This is our code:

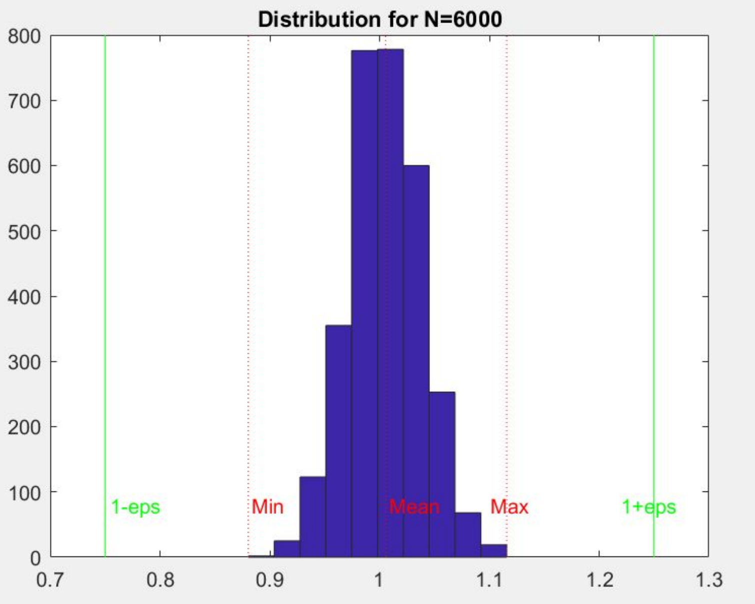
```
%% This program is for Ex 15.6
% To verify Johnson Lindenstrauss Lemma
close all
clear
BigN = [100;1000;3000;6000]; % Higher dimension
m = 3000; % Gaussian random vectors
eps = 0.125; % Distortion value
prob = 0.750;

for i=1:length(BigN)
      N=BigN(i);
      beta= -(log(1-prob))/log(m);
      % Using book’s JL lemma, minimum lower dimension
      n= ceil((2*beta*log(m)+2*log(m*(m-1)))/(eps^2/2-eps^3/3));
      R=rand(N,m);
      % Projection matrix
      A=randn(n,N);
      Proj=A./sqrt(n);

      % When Projected
      L=Proj*R;

      %Distortion
      for j=1:m-1
      Dis(j)= norm(L(:,1)-L(:,j+1))^2/norm(R(:,1)-R(:,j+1))^2;
      LB(j) = (1-eps);
      UB(j) = (1+eps);
      end

      mindis=min(Dis);
      maxdis=max(Dis);
      meandis=mean(Dis);

      figure
      hist(Dis)
      hold on
      vline(mindis,'r:','Min')
      vline(maxdis,'r:','Max')
      vline(meandis,'r:','Mean')
      vline(1-eps,'g-','1-eps')
      vline(1+eps,'g-','1+eps')
end
```

Results for the suggested (eps = 0.25, beta=0.75)





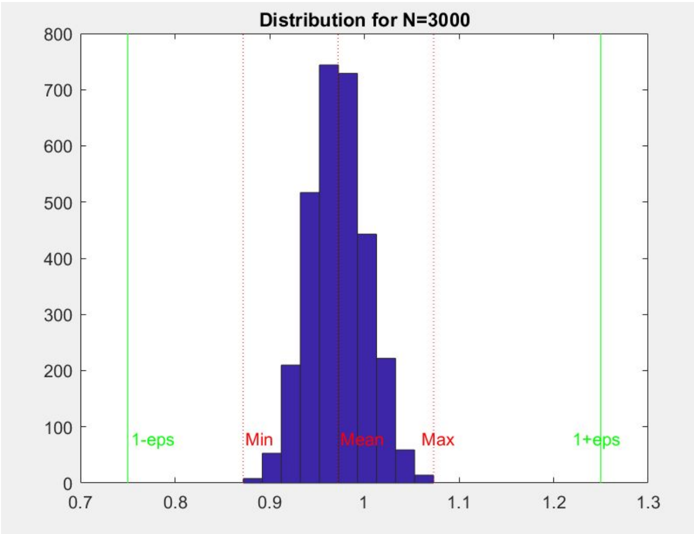


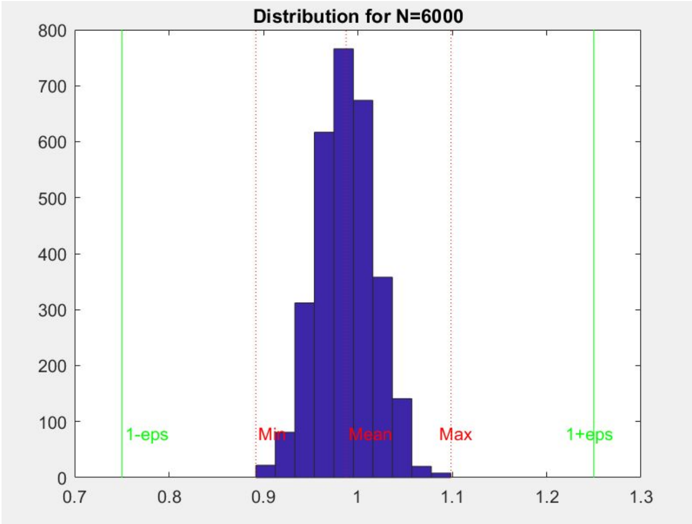


The dimension is only reduced for N>1337 as this is what is required by the specified tolerance. For N<1337 the data is 'projected' to a higher dimensional space and that the lengths are approximately only preserved is a useless result.

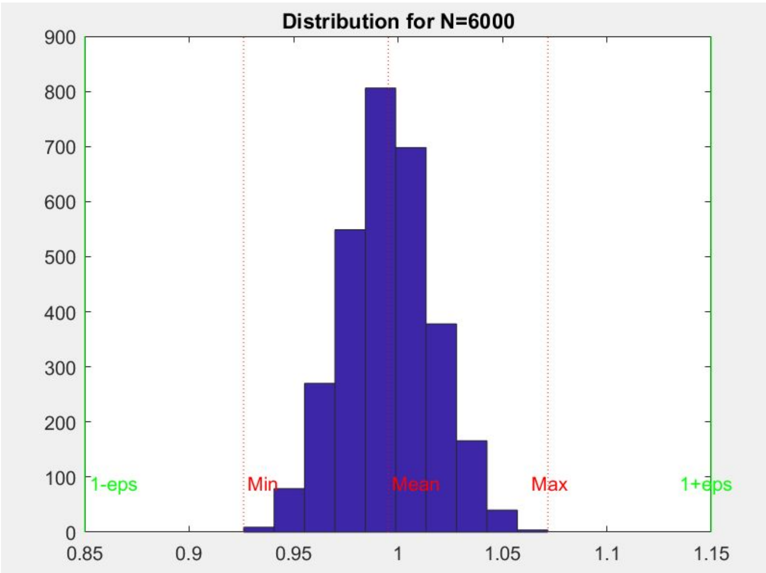
But for N>k that the distances are preserved well within the specified tolerance. We can see that the error bound given by the book is quite conservative.

Results for increased probability of success. (eps=0.25, beta=0.99)

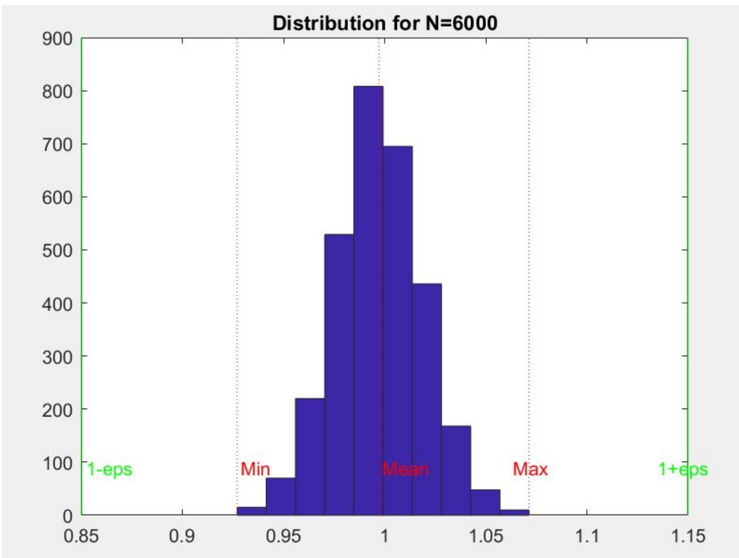




(eps=0.05, beta=0.5) leads to k= 27651
(eps=0.1, beta=0.5) leads to k= 7160
(eps=0.15, beta=0.5) leads to k=3300 which is pictured below for N=6000



Even when setting (eps=0.15, beta=0) k=3163) all of the distorted distances lie within 1+/-0.1. Showing that dimension error bound in the book is very loose and could be made sharper



Ex 15.7:

Ex. 15.7

Given, $m_i, i=1..N$ are sub-exponential, $a \in \mathbb{R}^N$, we have to show that

$$\mathbb{P}\left[\left|\sum_{i=1}^{N} a_i m_i\right| \geq t\right] \leq 2\min\left\{e^{-\frac{t^2}{c_1}}, c_2 e^{-\frac{t}{c_3}}\right\}$$

$$\underset{\substack{\downarrow \\ \text{Chernoff Bound}}}{\mathbb{P}\left[\sum a_i m_i \geq t\right] \leq} \min_{\lambda>0} \frac{\prod_{i=1}^{N} \mathbb{E}\left[e^{\lambda a_i m_i}\right]}{e^{\lambda t}} \underset{\substack{\downarrow \\ (15.6)}}{\leq} \min_{\lambda \leq \frac{1}{2\max(|a_i|\beta_i)}} e^{2\lambda^2 \sum_{i=1}^{N} a_i^2 \beta_i^2 - \lambda t}$$

$$\mathbb{E}\left[e^{\lambda a_i m_i}\right] \leq e^{2 a_i^2 \lambda^2 \beta_i^2} \quad \forall\, |a_i \lambda| \leq \frac{1}{2\beta_i}$$

Solving the minimization problem, $\lambda^* = \min\left\{\frac{t}{4\sum_{i=1}^{N} a_i^2 \beta_i^2}, \frac{1}{2\max(|a_i|\beta_i)}\right\}$

$$\Rightarrow \quad \mathbb{P}\left[\cancel{}\sum_{i=1}^{N} a_i m_i \geq t\right] \leq \min\left\{e^{-\frac{t^2}{8\sum_{i=1}^{N} a_i^2 \beta_i^2}}, e^{\frac{\sum a_i^2 \beta_i^2}{2\max(a_i^2 \beta_i^2)}} \cdot e^{-\frac{t}{2\max(|a_i|\beta_i)}}\right\}$$

$$c_1 = 8\sum_{i=1}^{N} a_i^2 \beta_i^2, \quad c_2 = \exp\left(\frac{\sum a_i^2 \beta_i^2}{2\max(a_i^2 \beta_i^2)}\right), \quad c_3 = 2\max(|a_i|\beta_i)$$

Using the union bound and symmetricity, we can say

$$\mathbb{P}\left[\left|\sum_{i=1}^{N} a_i m_i\right| \geq t\right] \leq 2\min\left\{e^{-\frac{t^2}{c_1}}, c_2 e^{-\frac{t}{c_3}}\right\}$$

For $a_i = \frac{1}{N}$, $i = 1, \ldots, N$,

$$\mathbb{P}\left[\left|\sum_{i=1}^{N} \frac{1}{N} m_i\right| \geq t\right] \leq 2\min\left\{e^{\left(\frac{-Nt^2}{\underbrace{8\sum_{i=1}^{N} \frac{\beta_i^2}{N}}_{c_1}}\right)}, \underbrace{e^{\frac{\sum \beta_i^2}{2\max \beta_i^2}}}_{c_2} \cdot e^{\frac{-Nt}{\underbrace{2\max(\beta_i)}_{c_3}}}\right\}$$