1 Algo 1: Sampling the Unit Ball

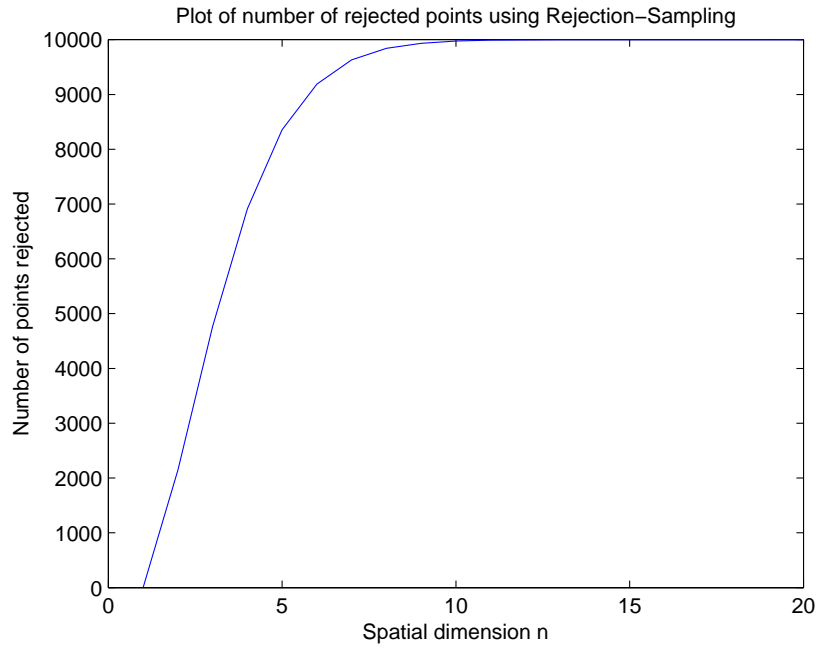**Q2. Plot the number of points rejected in $Algo1$ as function of n**

Figure 1: Plot of number of points rejected in Algo 1 as a function of spatial dimension

It is observed that as the dimensions increases, the number of points rejected increases exponentially. It becomes extremely difficult to obtain a point inside the Unit-Circle by the method of Rejection-Sampling.

By the method of construction, all the points lie within the Cube of length 2 units. However, we know that the volume of the unit-ball in n-dimensions varies according to the law:

$$V(B^n(1)) = \frac{\pi^{n/2}}{\Gamma(1+n/2)}$$

$$\lim_{n \to \infty} V(B^n(1)) = 0.$$

Thus, the volume of the unit-ball decreases with an increase in dimension of the space. The probability of randomly sampling a point inside the unit ball decreases in proportion to the ratio of the volumes of the Unit-Ball to the Cube of side 2-units, i.e.

$$\text{Prob(sample inside Unit Ball)} = \frac{V(B^n(1))}{V(\text{cube})} = \left(\frac{1}{2}\right)^n \frac{\pi^{n/2}}{\Gamma(1+n/2)}$$

2   Sampling the Unit Sphere

2.1   Experiments

Q3. Generate 10000 samples in $\mathcal{R}^n$ from a Gaussian distribution($\mu = 0$, $\sigma = 1$)
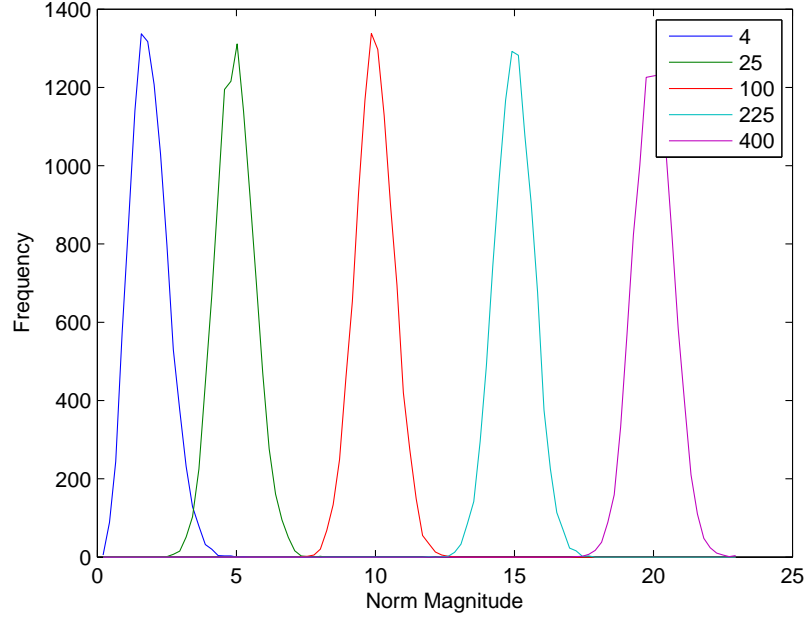

Figure 2: Q4. Plot of histogram of $||x||$ of points Gaussian(0,1) Distributed points in $\mathcal{R}^n$ with n=4,25,100,225,400

Q5. Conjecture on $||x||$, when x is Gaussian distributed in $\mathcal{R}^n$ What do you notice? Make a conjecture on the concentration of norm(x), when x is gaussian distributed in Rn.

We notice that the n-dimensional points obtained by sampling Gaussian Distribution($\mu = 0, \sigma^2 = 1$) are concentrated about a spherical shell of radius,$r = \sqrt{(n)}$. This behaviour is noted in the pattern of the horizontal parabolic curve of the variations of mean of the distribution with spatial dimension in Figure(3). Note the variance of the norms about the means is approximately constant at 0.5, suggesting that the Gaussian distribution is preserved.

2.2   Concentration of the Gaussian Measure

Let γ be the Gaussian Measure. For any measurable set $A \subset \mathcal{R}^n$, γ is defined by the integral over A,

$$\gamma(A) = (2\pi)^{-n/2} \int_A e^{-||x||^2/2} dx \tag{1}$$

Note that the integral is really computed using the density:

$$d\gamma(x) = d\gamma(x_1, x_2, \cdots x_n) = (2\pi)^{-n/2} e^{\frac{1}{2}\sum_{t=1}^n x_t^2} dx_1 dx_2 \cdots dx_n \tag{2}$$

Let $f(x)$ be a function defined on $\mathcal{R}^n$ taking values in $\mathcal{R}$, i.e. $f(x) : \mathcal{R}^n \to \mathcal{R}$.
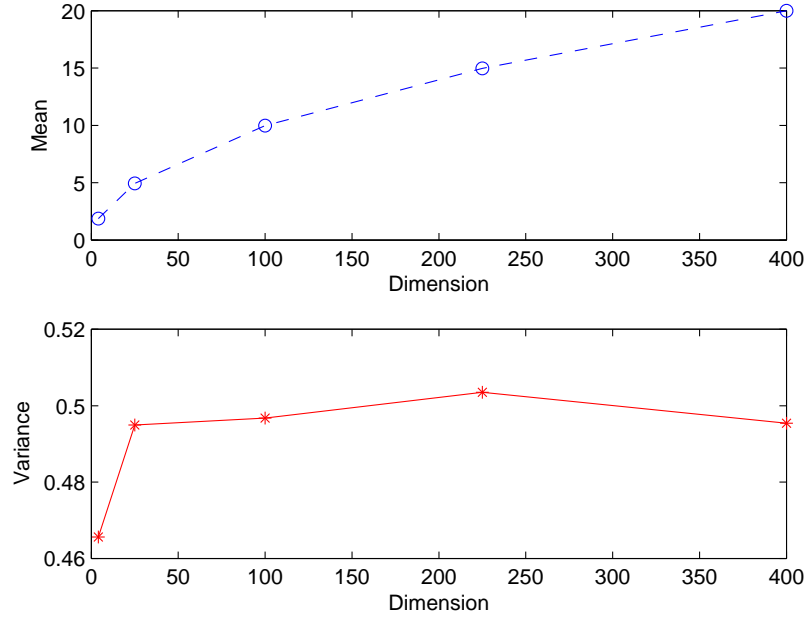
Figure 3: Plot of Mean and Variance of above distributions as a function of spatial dimension

Q6. Define the set $Q = \{x \in \mathcal{R}^n : f(x) \leq a, a \in \mathcal{R}\}$. We are required to prove:

$$\gamma\{Q\} \leq e^{a\lambda} \int_{\mathcal{R}^n} e^{-\lambda f(x)} d\gamma(x) \tag{3}$$

We know that $\forall x \in Q$,

$$f(x) \leq a \implies e^{-\lambda f(x)} \geq e^{-\lambda a}$$

Therefore, we have

$$\int_Q e^{-\lambda f(x)} d\gamma(x) \geq \int_Q e^{-\lambda a} d\gamma(x)$$
$$\geq e^{-a\lambda} \int_Q d\gamma(x)$$
$$\geq e^{-a\lambda} \gamma(Q).$$

Further, note that for $Q^c = \{x \in \mathcal{R}^n : f(x) > a\}$

$$\int_{\mathcal{R}^n} e^{-\lambda f(x)} d\gamma(x) = \int_Q e^{-\lambda f(x)} d\gamma(x) + \int_{Q^c} e^{-\lambda f(x)} d\gamma(x)$$

For the following result to hold true,

$$\int_{\mathcal{R}^n} e^{-\lambda f(x)} d\gamma(x) \geq \int_Q e^{-\lambda f(x)} d\gamma(x) \geq e^{-a\lambda} \gamma(Q) \tag{4}$$

we require that

$$\int_{Q^c} e^{-\lambda f(x)} d\gamma(x) \geq 0$$

1-3

Similar to the steps followed above, we find that

$$f(x) > a \implies -\lambda f(x) < -a\lambda \implies 0 < e^{-\lambda f(x)} < e^{-\lambda a}$$

$$0 < \int_{Q^c} e^{-\lambda f(x)} d\gamma(x) < \int_{Q^c} e^{-a\lambda} d\gamma(x) < e^{-a\lambda}\gamma(Q^c)$$

Thus, we have proved the validity of (4), to obtain

$$\gamma(Q) \le e^{a\lambda} \int_{\mathcal{R}^n} e^{-\lambda f(x)} d\gamma(x)$$

Q7.

Using the following substitution, $f(x) = ||x||^2/2$ and $a = (n - \delta)/2$, we get

$$Q = \{x \in \mathcal{R}^n : ||x||^2/2 \le (n - \delta)/2\}$$
$$= \{x \in \mathcal{R}^n : ||x||^2 \le (n - \delta)\} \tag{5}$$

Further, substituting into (3) we get

$$\gamma(Q) \le e^{\lambda(n-\delta)/2} \int_{\mathcal{R}^n} e^{-\lambda||x||^2/2} d\gamma(x)$$

$$\le e^{\lambda(n-\delta)/2} \int_{\mathcal{R}^n} e^{-\lambda||x||^2/2} \left((2\pi)^{-n/2} e^{||x||/2} dx_1 \cdots dx_n \right)$$

$$\le \frac{e^{\lambda(n-\delta)/2}}{(2\pi)^{n/2}} \int_{\mathcal{R}^n} e^{-(\lambda+1)||x||^2/2} dx$$

Q8. Using the substitution $y = x\sqrt{1 + \lambda}$, $\lambda = \delta/(n - \delta)$ and using the result

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathcal{R}^n} e^{-\frac{\lambda+1}{2}||x||^2} dx = (1 + \lambda)^{-n/2} \tag{6}$$

we proceed as follows

$$\lambda(n - \delta) = \delta \implies (1 + \lambda) = \frac{n}{n - \delta}$$

$$\gamma(Q) \le e^{\lambda(n-\delta)/2} \left(\frac{1}{(2\pi)^{n/2}} \int_{\mathcal{R}^n} e^{-(\lambda+1)||x||^2/2} dx \right)$$

$$\le e^{\delta/2}(1 + \lambda)^{-n/2}$$

$$\le e^{\delta/2} \left(\frac{n - \delta}{n} \right)^{n/2}$$

Q9. Let $\epsilon = \delta/n$, then we have $n - \delta = n(1 - \epsilon)$ and

$$\gamma(Q) \le e^{\delta/2} (\frac{n - \delta}{n})^{n/2}$$

$$\le e^{n\epsilon/2}(1 - \epsilon)^{n/2}$$

$$\log \gamma(Q) \le \frac{n}{2} \left(\log(1 - \epsilon) + \epsilon \right)$$

$$\le \frac{n}{2}(\frac{-\epsilon^2}{2})$$

where we use the result, $\log(1 - x) + x \leq -x^2/2$.

Q10. Using similar methods, we can derive the following results

$$\gamma \left\{ x \in \mathcal{R}^n : ||x||^2 \geq \frac{n}{1 - \epsilon} \right\} \leq e^{-n\epsilon^2/4} \tag{7}$$

and

$$\gamma \left\{ x \in \mathcal{R}^n : ||x||^2 \geq n(1 + \epsilon) \right\} \leq e^{-n\epsilon^2/8} \tag{8}$$

From(7) we note that

$$\gamma \left\{ x \in \mathcal{R}^n : \left(\frac{||x||^2}{n} - 1 \right) \leq -\epsilon \right\} \leq e^{-n\epsilon^2/4}$$

and from (8) we have

$$\gamma \left\{ x \in \mathcal{R}^n : \left(\frac{||x||^2}{n} - 1 \right) \geq \epsilon \right\} \leq e^{-n\epsilon^2/8} \tag{9}$$

Combining the above two results, we obtain

$$\gamma \left\{ x \in \mathcal{R}^n : \left| \frac{||x||^2}{n} - 1 \right| \geq \epsilon \right\} \leq 2 * \text{Max}(e^{-n\epsilon^2/8}, e^{-n\epsilon^2/4}) \tag{10}$$

$$\leq 2e^{-n\epsilon^2/8} \tag{11}$$

Q11. Explain why this result shows that the Gaussian measure is concentrated on the sphere of radius of $\sqrt{n}$ with a decay of $e^{-\epsilon^2/8}$.

Note that by definition, i.e.

$$Q = \left\{ x \in \mathcal{R}^n : \left| \frac{||x||^2}{n} - 1 \right| \geq \epsilon \right\}$$

the set Q consists of all points in n-dimensional space with the square of ratio of its norm to a circle of radius $r = \sqrt{n}$ being separate from Unity(1) by a margin of ϵ. The Gaussian Measure of set Q, $\gamma(Q)$ is exponentially decreasing in ϵ, as suggested by (11), suggesting that the Gaussian measure in n-dimensional space is concentrated on a sphere of radius $\sqrt{n}$.

2.3   Algo 2: Generate Uniform samples on sphere $S^{n-1}(r)$

$N = 10,000$ points on the sphere $S^{n-1}(1)$ are sampled using the given algorithm. The projections of the points on the x_1 axis is derived. The histogram of the projections in Figure(4), along with the mean and variance of the distributions in Figure(5) are obtained.

We observe that the projections of the points are distributed as a Gaussian with exponentially decreasing variance as the number of spatial dimensions increases. This implies that as the spatial dimension increases, with very high probability the projection of an arbitrary point in the Unit-Ball shall equal Zero.
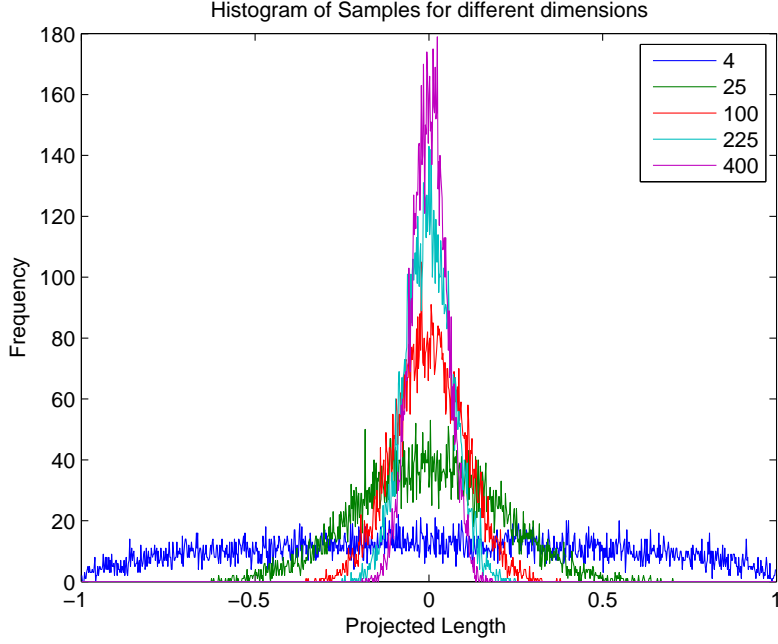
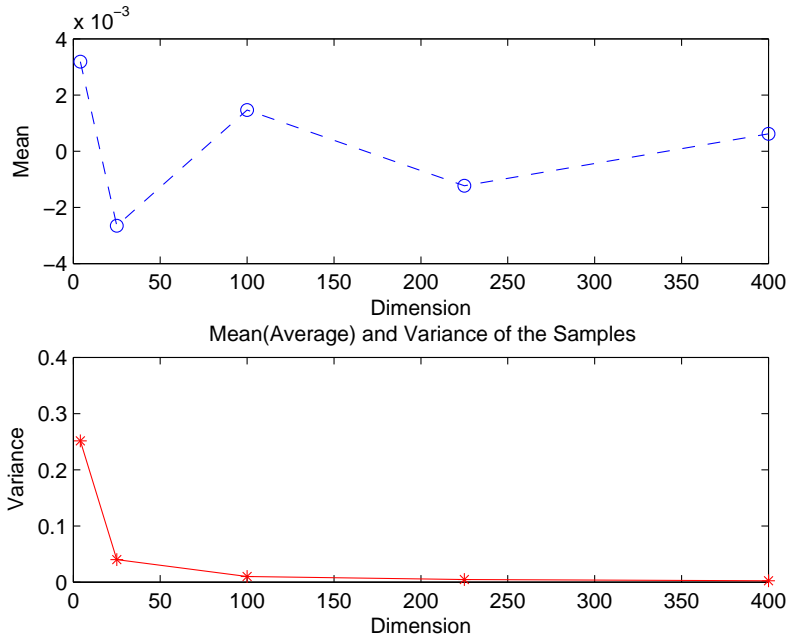Figure 4: Q12. Histogram of projections of points sampled on Sphere $S^{n-1}(1)$ for n = 4, 25, 100, 225, 400

Figure 5: Q13. Plot of Mean and Variance of above distributions as a function of n

Q14. Determine $\epsilon(n)$ **such that** $0.99N$ **points in the slab** $x = (x_1, ..., x_n) \in S^{n-1}(1) : -\epsilon(n) \leq x_1 \leq \epsilon(n)$.

In Figure(6), we try to determine the thickness of the slab, that shall contain 0.99N of the generated points. We expect that as the number of dimensions increases, the thickness

of the slab decreases exponentially as more and more points get projected closer to Zero(0). The related pattern is obtained by plotting the cummulative frequency of points within a given thickness of slab in Figure(7). The curves with steepest slope corresponds to sample points in the higher n-dimensional space.
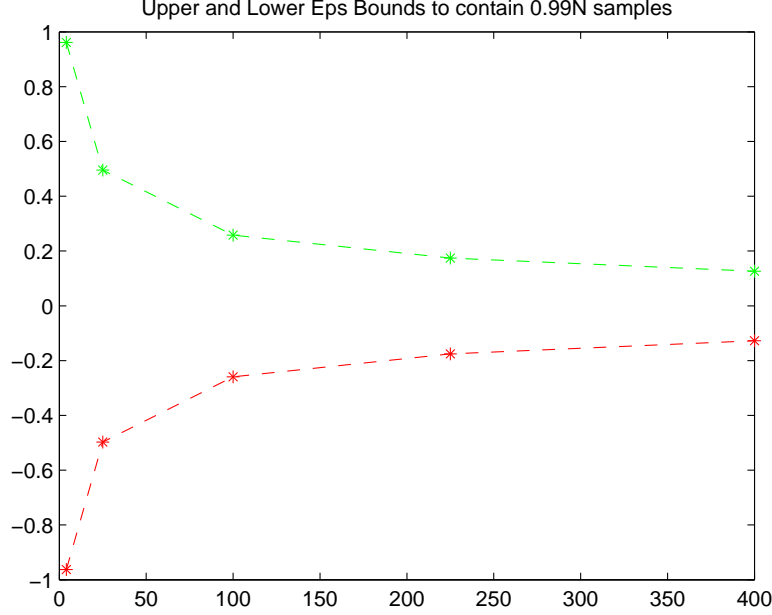

Figure 6: Plot of $\epsilon(n)$ as a function of spatial dimension n.

Q15. Do you think the choice of the axis is important for the results obtained? Justify your answer.

It is important to note that, the choice of the direction/axis of projection determines the magnitude of the obtained value of projection. It is known from theory that if a direction along a diagonal of the cube is chosen, then the projection of the points in the sphere upon the diagonal shall have a length of Unity(1), while the projections along the orthogonal axis shall have length Zero(0).

Therefore, the choice of projecting the points to either of the orthogonal axis $x_1, x_2, \cdots x_n$ are equivalent, as the projections shall concentrate about Origin(Zero). However, the choice of direction along a diagonal of the Cube shall have projections that concentrate around Unity(One).

3 Algo 3: Uniform sampling in the Ball, $B^n(\sqrt{}(n))$

$N = 10,000$ points inside Unit Sphere $B^n(\sqrt{n})$ are generated by sampling with Uniform Measure. Projections of the points on the x_1 axis is computed. The histogram of the projections and the mean/variance of the distribution of these projections is computed and plotted for $n = 4, 25, 100, 225, 400$.

In an n-dimensional space, we expect the ball of radius $r = \sqrt{n}$ to have Unit-Volume i.e. $Volume(B^n(\sqrt{n})) = 1$. It is observed that, all the projections are approximately Gaussian Distributed($\mu = 0, \sigma^2 = 1$) independent of the dimension of the space(n) considered.

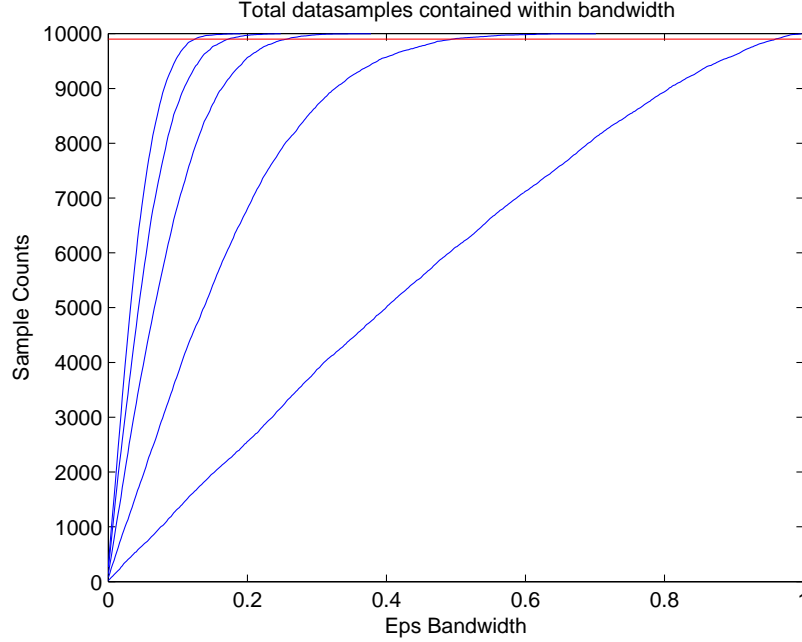**Q19. What curve would you expect to find when measuring the relative volume**

Figure 7: Plot of sample-counts-vs- $\epsilon(n)$ to determine bandwidth for different spatial dimensions, n.

of a slab of thickness 1/2 in the $B^n(\sqrt{n})$? Further, from the nature of the histogram, we expect that the relative volume $w(n)$ to remain approximately constant. This is verified by the plot of the relative volume obtained in Figure(10).

Q20. Is the choice of axis of any consequence for the above results? The choice of the axis among the orthogonal axes $x_1, x_2, \cdots x_n$ is not important for the results in the above questions. However, as earlier, the choice of axes along diagonal of a cube with edges along the orthogonal axis, would indicate different results.

Q21. Plot histogram of the distances of the points from the origin The histogram of the distances of the points sampled in $B^n(\sqrt{n})$ is presented in Figure(11). We note that the points sampled unifromly inside a n-dimensional ball of radius $r = \sqrt{n}$ are concentrated in a thin shell.

The apparent paradox, is that though the projections of sample-points uniformly sampled inside a sphere are approximately gaussian distributed about Origin, the points are concentrated within a thin-shell about the radius of the Sphere. The n-dimensional solid-sphere has "EMPTY SPACE" inside it and consists of a "THIN SHELL-CRUST" harbouring the entire volume of the sphere.

Q18.

Q20. Do you think the choice of axis is important for the above results? Justify your answer!

Q21.

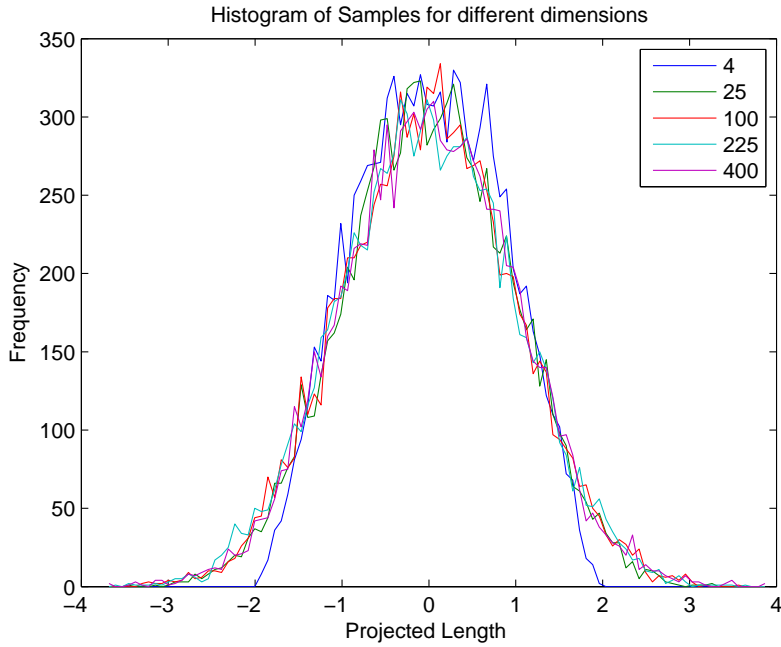# 4 Empirical distribution of eigenvalues and singular values

Q23.

Figure 8: Q17. Histogram of projection of uniformly sampled points inside Unit Sphere upon axis x_1 with varying spatial dimension n.
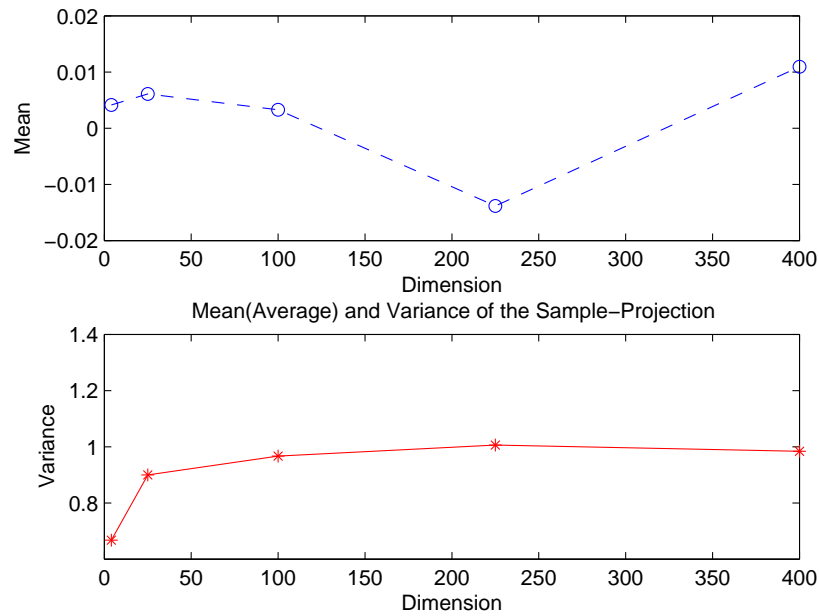

Figure 9: Q18. Plot of mean and variance of the distributions of projection of uniformly sampled points inside Unit Sphere upon axis x_1 as a function of spatial-dimension n.

Q24. Find the line of best-fit to guess the dependency of λ and λ with spatial-dimension n. The line of best-fit was computed for the plots of λ-vs-$log(n)$.
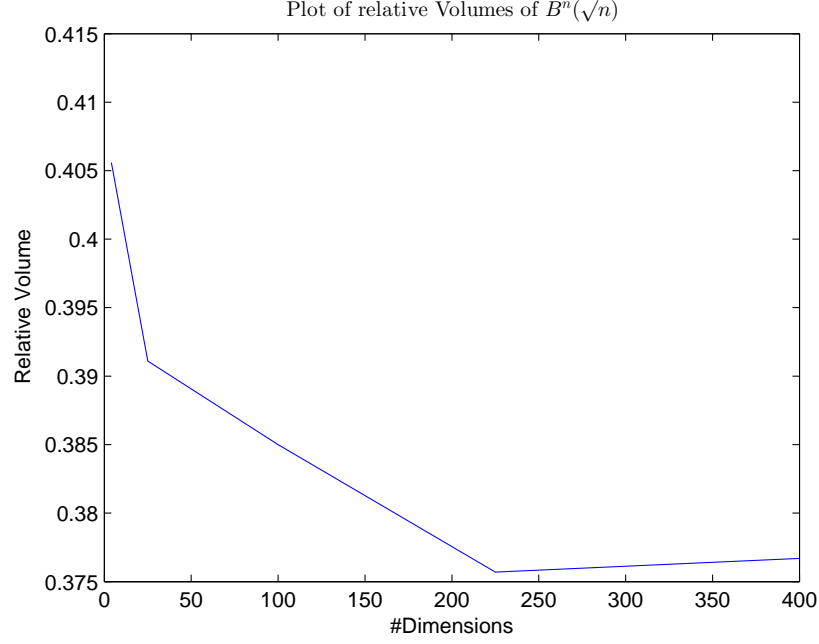
Figure 10: Q.19. Plot of Relative Volume $w(n)$ measured in a slab of thickness $1/2$ in $B^n(\sqrt{n})$ as a function of spatial dimension n.
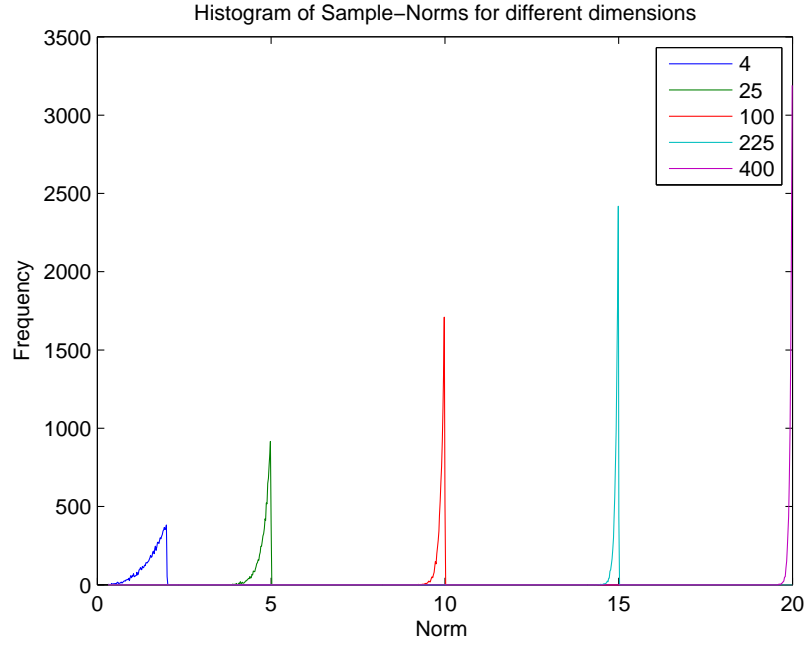
Figure 11: Q.21. Histogram of $||x||$ for uniformly sampled points in $B^n(\sqrt{n})$ for spatial-dimension $n = 1, ..., 400$.

For the Wigner Gaussian Matrices we obtained the straight-line equations as

$$\lambda_{max} = 12.0643 * log(n) - 26.6017$$
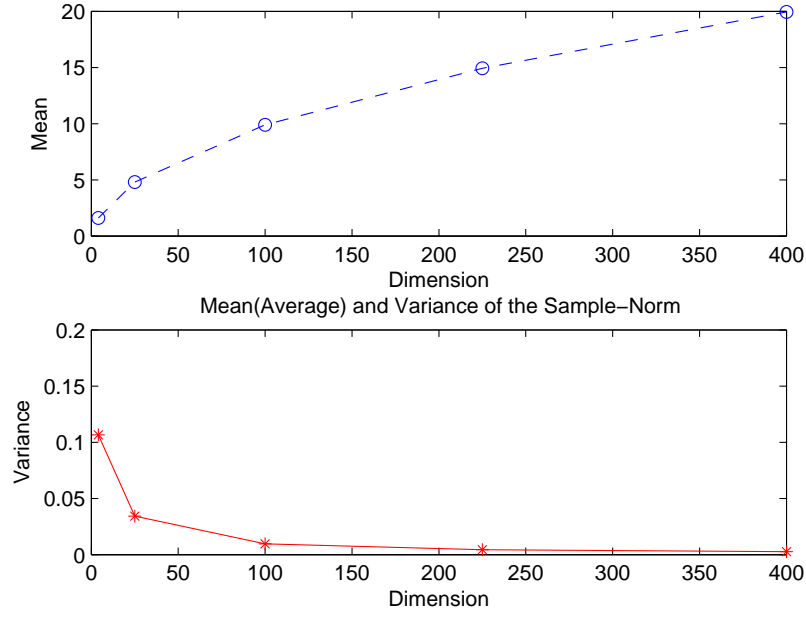$$\lambda_{min} = -11.9874 * log(n) + 26.6017$$

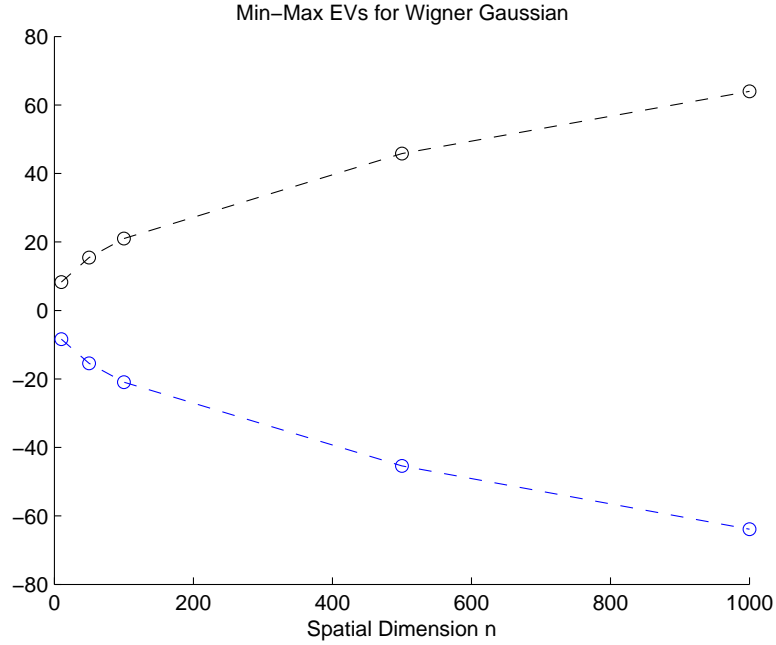Figure 12: Q.21. Mean and variance of above distributions of $||x||$ for uniformly sampled points in $B^n(\sqrt{n})$

Figure 13: Q23. Wigner Gaussian: Scatter plot of λ_{min} and λ_{max} over 100 realizations for n=10, 50, 100, 500, 1000.

For the Wigner Bernoulli Ensembles, we obtained the straight-line equations as

$$\lambda_{max} = 12.4254 * log(n) - 29.1955$$
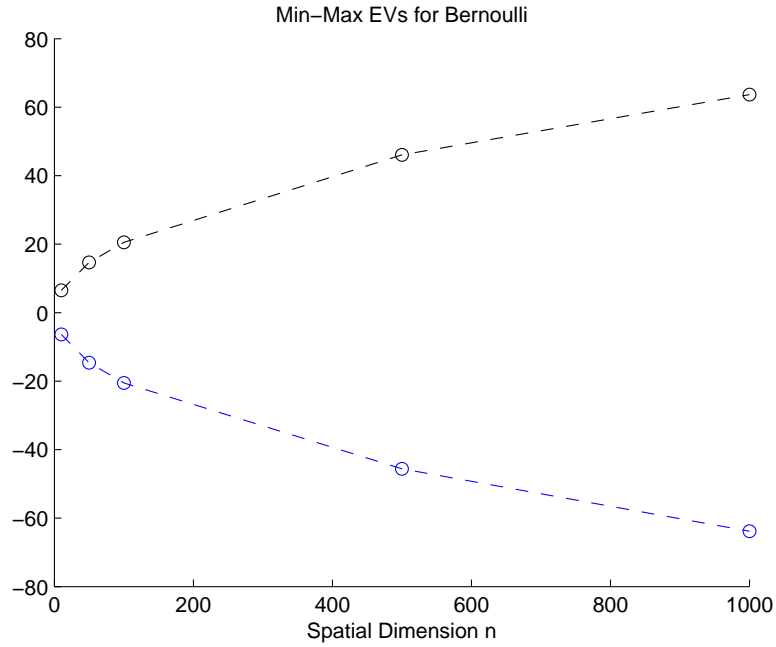$$\lambda_{min} = -12.4371 * log(n) + 29.3655$$

Figure 14: Q23. Bernoulli Ensemble: Scatter plot of λ_{min} and λ_{max} over 100 realizations for n=10, 50, 100, 500, 1000.
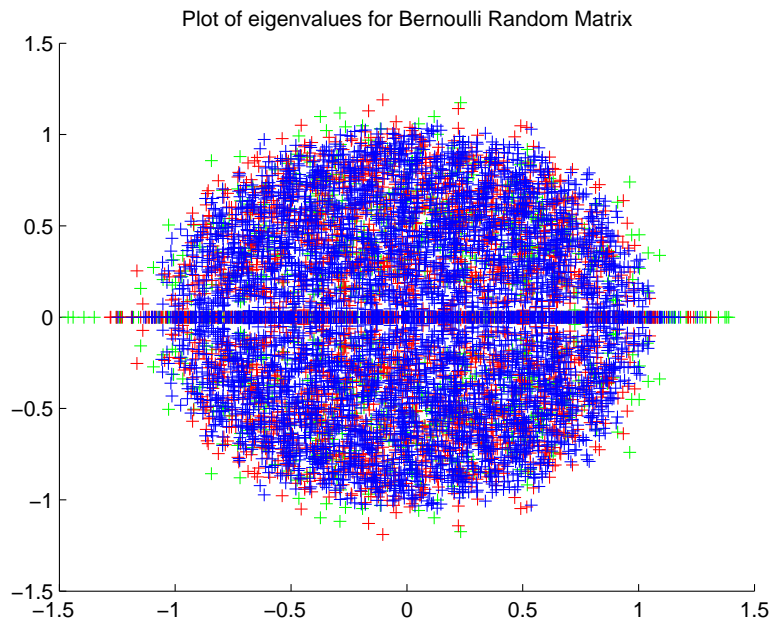
Q26.

Figure 15: Q26. Scatter-plot of normalized eigenvalues of asymmetric Bernoulli random matrices with n=10, 50, 100, 500, 1000 over 100 realizations
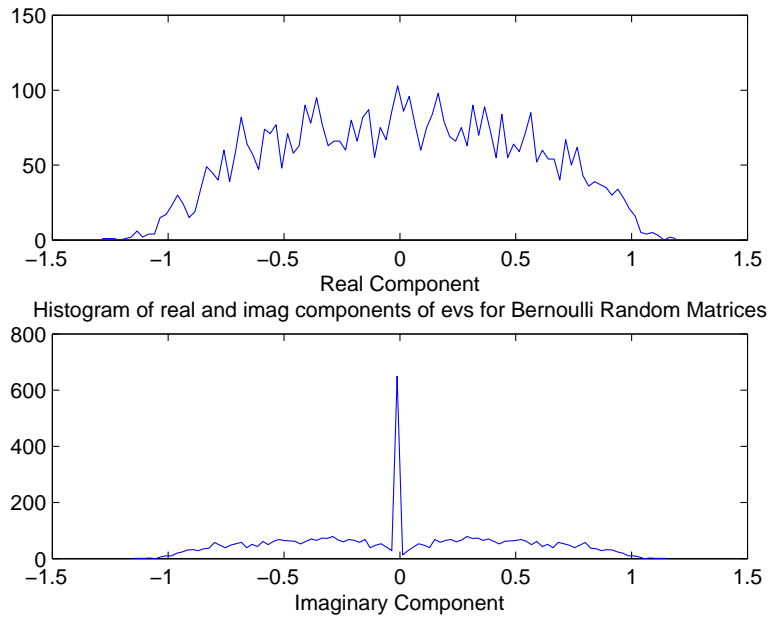
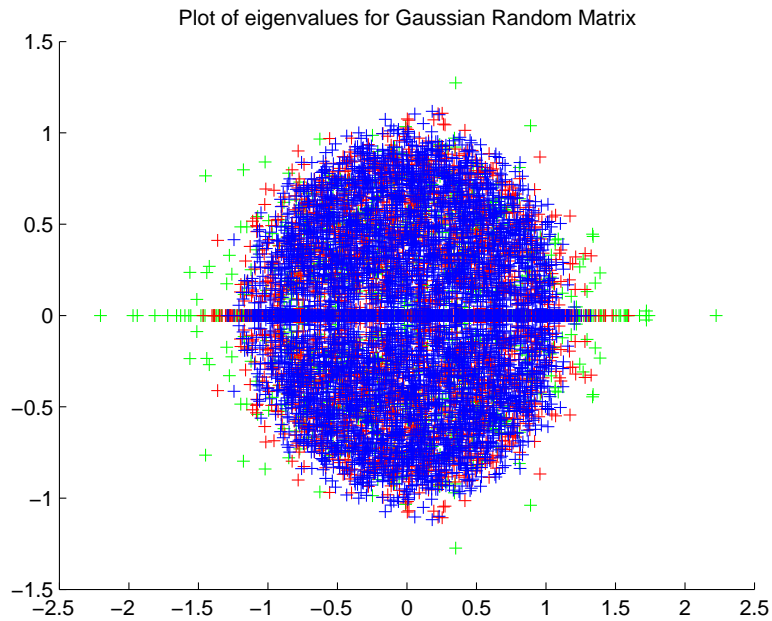Figure 16: Q26. Histogram of real and imaginary components of the above normalized eigenvalues

Figure 17: Q26. Scatter-plot of normalized eigenvalues of asymmetric Gaussian random matrices with n=10, 50, 100, 500, 1000 over 100 realizations

Q27. Conjecture on limit distribution of spectrum of random square matrices
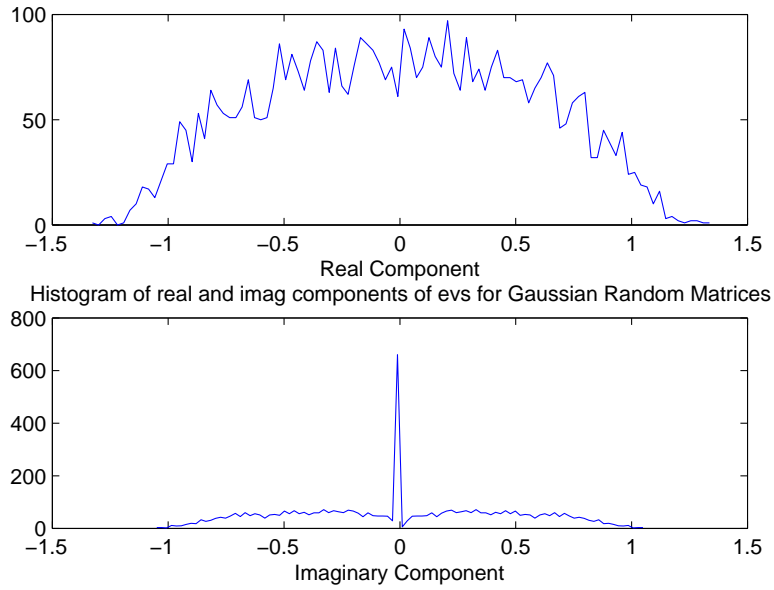
Figure 18: Q26. Histogram of real and imaginary components of the above normalized eigen-values

with independent entries which are non-symmetric

The spectruc of random square matrices with independent entries is lies within a circle in the complex plane with radius $r = 1.5$. Most of the eigenvalues are concentrated about the Real-Axis. However, they are symmetrically distributed about the origin, with maximum number of eigenvalues being concentrated nearer to the origin than farther away from it. The plots of the real and imaginary components of the eigenvalues for the random matrices are shown in the figures.