Chapter 10

High dimension

10.1 Introduction: Volume of the unit ball in d dimensions

In this chapter we will deal with the geometry of high dimensions, specifically we will touch upon the volume of unit balls and how to sample from them and from unit spheres. Generating points from a unit ball or sphere is very useful thing in applications and simulation.

Definition 10.1. *Given a radius $r > 0$ we define the d-***dimensional ball** *as the set*

$$B_r(x) := \{y \in \mathbb{R}^d : |x - y| < r\}.$$

*We also denote the d-***dimensional sphere** *as the set*

$$S_r(x) := \{y \in \mathbb{R}^d : |x - y| = r\}.$$

Whenever $r = 1$ we call $B_1(x), S_1(x)$ **unit ball** *and* **unit sphere** *respectively. If $x = 0$ we omit it from the notation, and use $B_r = B_r(0)$ and $S_r = S_r(0)$.*

Remark 10.2. *In this chapter we will be using the volume of sets in $\mathbb{R}^d$, but how do we define the volume? It is simply as follows*

$$|E| = \int_E dx = \int \mathbb{1}_E dx = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_E(x_1, \ldots, x_d) dx_1 \ldots dx_d.$$

But how does the volume of a unit ball change with dimension? Intuitively you would perhaps say that it does not change, so lets use the law of large numbers to convince you otherwise.

To start, let us define the spherical Gaussian

Model 10.3. *A continuous $\mathbb{R}^d$ valued random variable Z with density function*

$$f(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}|x|^2\right), \quad x \in \mathbb{R}^d$$

is called a **spherical Gaussian**. *In short, each coordinate is a standard Gaussian and are independent of eachother.*

Let us also define the normalized Gaussian

Model 10.4. *Let Z be a spherical Gaussian in $\mathbb{R}^d$ and consider $Y = (2\pi)^{-1/2}Z$ then Y is called a* **normalized Gaussian**, *the density is*

$$f(x) = \exp\left(-\pi|x|^2\right), \quad x \in \mathbb{R}^d.$$

Lemma 10.5. *Let $d > 4\pi$ then for $B_1 \subset \mathbb{R}^d$ there exists a constant $C > 1$ that does not depend on dimension such that*

$$|B_1| \leq \frac{C}{d}.$$

Proof. Let $Z \in \mathbb{R}^d$ be a normalized Gaussian, then we have that the density of Z, f, satisfies

$$\begin{aligned} &f(0) = 1 \\ &f(z) \geq e^{-\pi}, \quad z \in B_1. \end{aligned}$$

Now for the normalized Gaussian each component of $Z = (Z_1, \ldots, Z_d)$ are i.i.d. and they all are Gaussians with variance $(2\pi)^{-1/2}$, this means that $|Z|^2 = \sum_i |Z_i|^2$ is now a sum of independent r.v.s and if we use Proposition 3.2

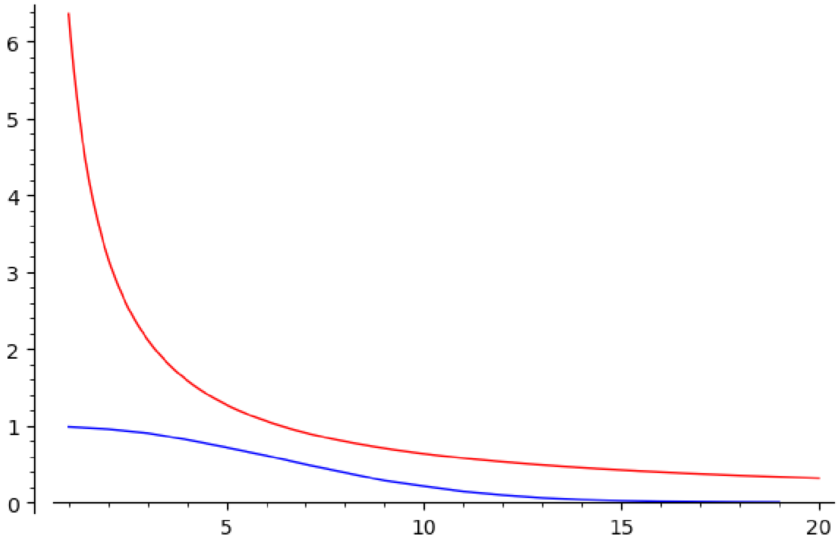
$$\mathbb{P}\left(\left||Z|^2 - \mathbb{E}\left[|Z|^2\right]\right| \geq \epsilon\right) \leq \frac{\operatorname{Var}(|Z|^2)}{\epsilon^2} = \frac{d \operatorname{Var}(|Z_1|^2)}{\epsilon^2} = \frac{cd}{\epsilon^2} \tag{10.1}$$

in the second to last step we used independence and in the last step we used that $\operatorname{Var}(|Z_1|^2) = c$ is a number that we can compute but we will skip it here. However, we also know that $\mathbb{E}\left[|Z|^2\right] = \frac{d}{2\pi}$, so if we choose $\epsilon = (d - 2\pi)/2\pi$ we get (f is the density for Z)

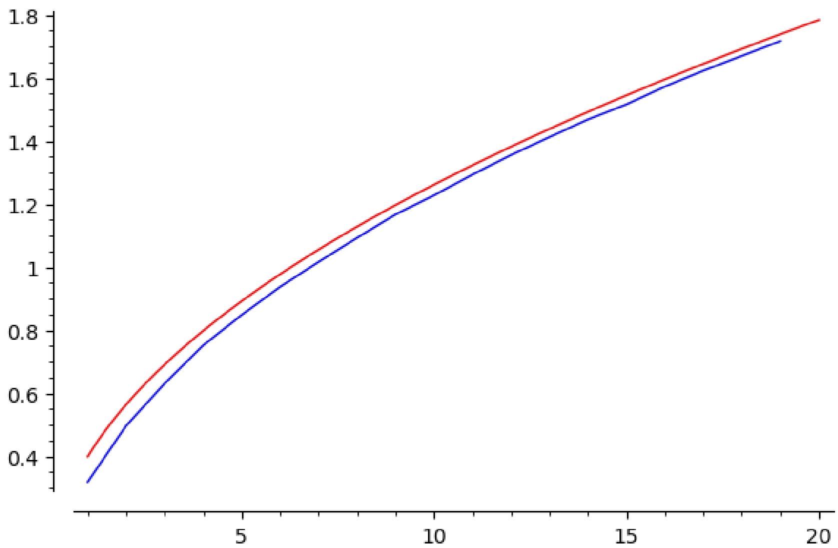
$$\begin{aligned} \frac{|B_1|}{e^\pi} &\leq \int_{B_1} f(z)dz = \mathbb{P}(Z \in B_1) \\ &\leq \mathbb{P}\left(|Z|^2 \geq \epsilon + \mathbb{E}\left[|Z|^2\right] \text{ or } \mathbb{E}\left[|Z|^2\right] - \epsilon \geq |Z|^2\right) \\ &= \mathbb{P}\left(\left||Z|^2 - \mathbb{E}\left[|Z|^2\right]\right| \geq \epsilon\right) \\ &\leq \frac{cd}{((d-2\pi)/(2\pi))^2} \leq \frac{C}{d} \end{aligned}$$

for some constant $C > 1$. $\square$

Thus the volume of the unit ball will decrease with dimension.



In the picture you can see the blue curve being the estimated probability of a Gaussian landing inside the unit ball for different dimensions, while the red curve denotes the upper bound given by Lemma 10.5. We did a fairly poor job at capturing the behavior, the actual volume seems to be much smaller than our estimate. However, the estimate (10.1) also suggests that $|Z|$ should concentrate around $\sqrt{\frac{d}{2\pi}}$, below you can see the plots of the estimated $|Z|$ vs the expected.



This seems fairly spot on, interesting!

Exercise 10.6. *The proof above used the concentration inequality (Chebyshev), which is a fairly weak one. Can you improve on the estimate above using another concentration inequality? Do this before you read on...*

10.2 The geometry of high dimension

The scaling property of volume. Lets say we have a cube centered at the origin, namely the cube can be written as $Q = [-l, l]^d$ where d is the dimension, the volume is the product of the side-lengths and thus $|Q| = (2l)^d$. Scaling each side of the cube by $(1-\epsilon)$ where ϵ is a small number gives us that the volume also scales with $(1-\epsilon)^d$, this gives us the formula

$$|(1-\epsilon)Q| = (1-\epsilon)^d |Q|$$

Lets divide this equation by the volume of Q, we get

$$\frac{|(1-\epsilon)Q|}{|Q|} = (1-\epsilon)^d \to 0$$

as $d \to \infty$. The conclusion is that most of the volume is located close to the surface of the cube.

Lemma 10.7. *Let $E \subset \mathbb{R}^d$ and let $\epsilon \in (0,1]$, then*

$$(1-\epsilon)^d |E| = |(1-\epsilon)E|$$

where $(1-\epsilon)E := \{(1-\epsilon)x : x \in E\}$.

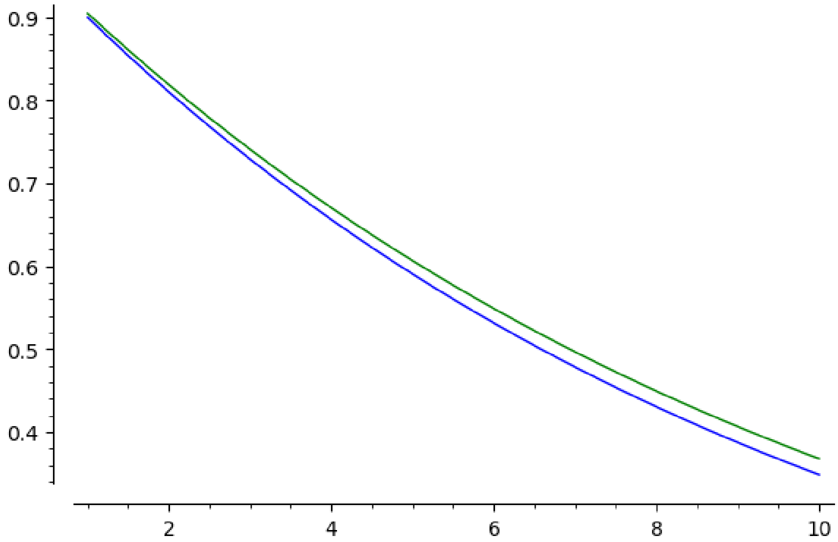
Proof. By the change of variables formula ($y_1 = (1-\epsilon)x_1$ we get $dy_1 = (1-\epsilon)dx_1$) and our area becomes

$$\begin{aligned}
|E| &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_E(x_1, \ldots, x_d) dx_1 \ldots dx_d \\
&= \frac{1}{(1-\epsilon)^d} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_E(y_1/(1-\epsilon), \ldots, y_d/(1-\epsilon)) dy_1 \ldots dy_d \\
&= \frac{1}{(1-\epsilon)^d} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_{(1-\epsilon)E}(y_1, \ldots, y_d) dy_1 \ldots dy_d \\
&= \frac{1}{(1-\epsilon)^d} |(1-\epsilon)E|
\end{aligned}$$

which is what we wanted to prove. □

[269]:
```
P=plot((1-0.1)^x,1,10)
P+=plot(exp(-0.1*x),1,10,color='green')
P.show()
```

[269]:



Based on the above, we can choose $\epsilon = 1/d$ which gives us that most of the volume is contained in the annulus below.

10.3 Properties of the unit ball

Let us now do one of the main computations of this chapter, namely the volume of the unit ball is computed exactly. The computation is a bit complicated but follows the following simple structure:

1. Write the volume of the unit ball as an integral of the constant function 1 over the unit ball
2. Use a radial coordinate system to rewrite that integral so that we get the integral over the surface of a unit ball instead.
3. Compute the integral of the Gaussian kernel in two ways, one using the fact that $\exp(|x|^2) = \exp(|x_1|^2)\exp(|x_2|^2)\ldots\exp(|x_d|^2$, the second one using radial coordinates
4. The radial part of the Gaussian integral gives rise to the Gamma function, which is a generalization of the factorial, the spherical part is just the area of the unit sphere (which is the one we are after).

Theorem 10.8. *The volume of the unit ball in d dimensions is*

$$|B_1| = \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}$$

where Γ *is the aptly named Gamma-function. If* k *is an integer then* $\Gamma(k) = (k-1)!$*, which gives us for even dimensions that*

$$|B_1| = \frac{2\pi^{\frac{d}{2}}}{d(\frac{d}{2}-1)!}.$$

Proof. We begin by first writing down what we want to compute

$$|B_1| = \int_{B_1} dx$$

Step 2: Surface integral

$$\int_{B_1} dx = \int_{S^d} \int_0^1 \left|\frac{dx}{dr}\right| dr d\Omega$$

where $\left|\frac{dx}{dr}\right|$ is the Jacobian of the change of variables and $d\Omega$ is the surface element on the unit sphere S_1 (think of the area of a tiny square on the surface of a ball, in 3d we can think of the longitude and latitue coordinates).

$$\left|\frac{dx}{dr}\right| = r^{d-1}$$

The conclusion is that

$$\int_{B_1} dx = \int_{S_1} \int_0^1 \left|\frac{dx}{dr}\right| dr d\Omega = \frac{|S_1|}{d}.$$

In the above we used $|S_1| := \int_{S_1} d\Omega$.

Step 3a: Gaussian kernel trick (repeated integrals)

This is where we use the Gaussian kernel trick, first note that

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

The standard normal random variable has a normalizing factor which is $1/\sqrt{\pi}$.

[273]:
```
x = var('x')
integrate(exp(-x^2),x,-infinity,infinity)
```

[273]:
```
sqrt(pi)
```

$$\int_{\mathbb{R}^d} e^{-|x|^2} dx = \int_{\mathbb{R}^d} \prod_{i=1}^{d} e^{-x_i^2} dx = \prod_{i=1}^{d} \int_{\mathbb{R}^d} e^{-x_i^2} dx_i = \pi^{d/2}$$

Step 3b: Gaussian kernel trick (spherical coordinates)

Let us compute the same integral again, but this time using spherical coordinates

$$\int_{\mathbb{R}^d} e^{-|x|^2}dx = \int_{S_1}\int_0^\infty e^{-r^2}r^{d-1}drd\Omega = |S_1|\int_0^\infty e^{-r^2}r^{d-1}dr$$

now doing the change of variables $t = r^2$ we get $dt = 2rdr$ and thus

$$\int_0^\infty e^{-r^2}r^{d-1}dr = \int_0^\infty e^{-t}t^{\frac{d-1}{2}}\frac{1}{2\sqrt{t}}dt = \frac{1}{2}\int_0^\infty e^{-t}t^{\frac{d}{2}-1}dt = \frac{1}{2}\Gamma\left(\frac{d}{2}\right)$$

In conclusion assembling the previous step with this

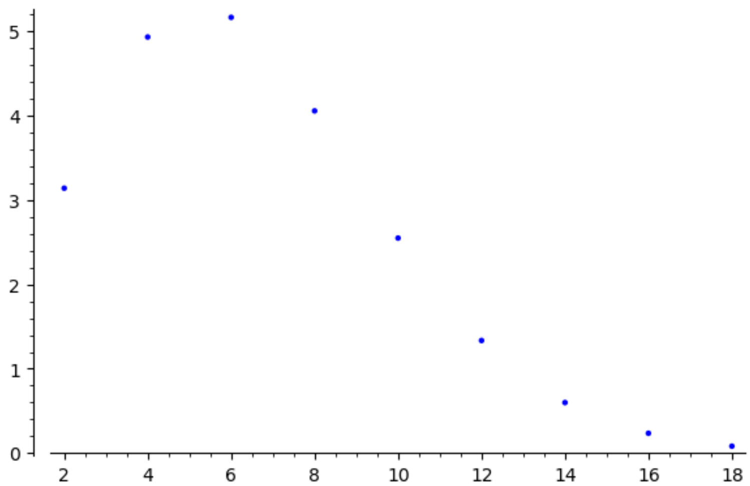
$$\pi^{d/2} = \int e^{-|x|^2}dx = |S_1|\frac{1}{2}\Gamma\left(\frac{d}{2}\right).$$

Step 4: Assembly time

Assemling step 1 and 2 together with the line above we get

$$|B_1| = |S_1|/d = \frac{\pi^{d/2}}{\frac{1}{2}\Gamma\left(\frac{d}{2}\right)d}$$

which completes the proof. $\square$



10.4 Uniform at random from a ball and sphere

Oftentimes we want to work with what is denoted as uniform at random from the unit sphere or the unit ball. Let us define these

Model 10.9. *We say that an $\mathbb{R}^d$ valued random variable Z is* **uniform at random from the unit sphere** *if $Z \in S_1$ and for any A we have*

$$\mathbb{P}(Z \in A) = \frac{1}{|S_1|} \int_{S_1} \mathbb{1}_A(\theta) d\Omega(\theta)$$

where the integral above is the surface integral on the sphere, here $d\Omega$ is the surface element on S_1. We denote this as $Z \sim \text{Uniform}(S_1)$.

Remark 10.10. *We havent really defined what a surface element is, but the heuristic understanding is enough for now.*

This definition is easier to grasp

Model 10.11. *We say that an $\mathbb{R}^d$ valued random variable Z is* **uniform at random from the unit ball** *if $Z \in B_1$ and for any A we have*

$$\mathbb{P}(Z \in A) = \frac{1}{|B_1|} \int_{B_1} \mathbb{1}_A(z) dz = \frac{|A \cap B_1|}{|B_1|}.$$

In short, the probability of landing inside $A \cap B_1$ is given by the proportion of the volume it makes up out of B_1. We say $Z \sim \text{Uniform}(B_1)$.

How about generation? Let us start with 2 dimensions.

10.4.1 Generating points uniformly at random from a circle

Lets say that we want to generate a uniformly at random variable on the unit circle. One suggestion would be to generate two coordinates X and Y i.i.d. from Uniform$(-1, 1)$ and then projecting (X, Y) onto the unit circle. However, this results in the picture below, when we plot the distribution of angles.

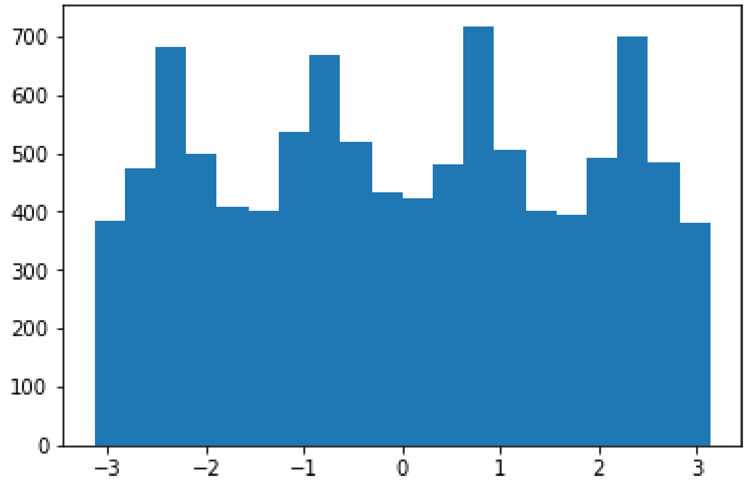
Lets take a look, denote the projection π, i.e.

$$\pi(x, y) = (X/\sqrt{X^2 + Y^2}, Y/\sqrt{X^2 + Y^2}),$$

then the density of the angle

$$f_{\angle \pi(X,Y)}(\theta) = c_0 \int_0^\infty p_{(X,Y)}(t \cos(\theta), t \sin(\theta)) dt \tag{10.2}$$

for some constant c_0. Since $p_{(X,Y)}(x, y)$ is constant (uniform distribution) the above basically measures the length of the line starting from the origin $(0, 0)$ and stretches out in direction $(\cos(\theta), \sin(\theta))$ and reaches the edge of the unit square $[-1, 1]^2$. This is why we see four peaks in the plot below, one for each corner.



Remark 10.12. *From (10.2) we see that if $p_{(X,Y)}$ is rotationally symmetric, i.e. $p_{(X,Y)}(t\cos(\theta), t\sin(\theta)) = p(t)$ for some p then the distribution over angles become uniform.*

This warrants the following definition

Definition 10.13. *We say that a function $f : \mathbb{R}^d \to \mathbb{R}$ is* **rotationally symmetric** *if there exists a function $g : \mathbb{R} \to \mathbb{R}$ such that*

$$f(x) = g(|x|)$$

for all $x \in \mathbb{R}^d$.

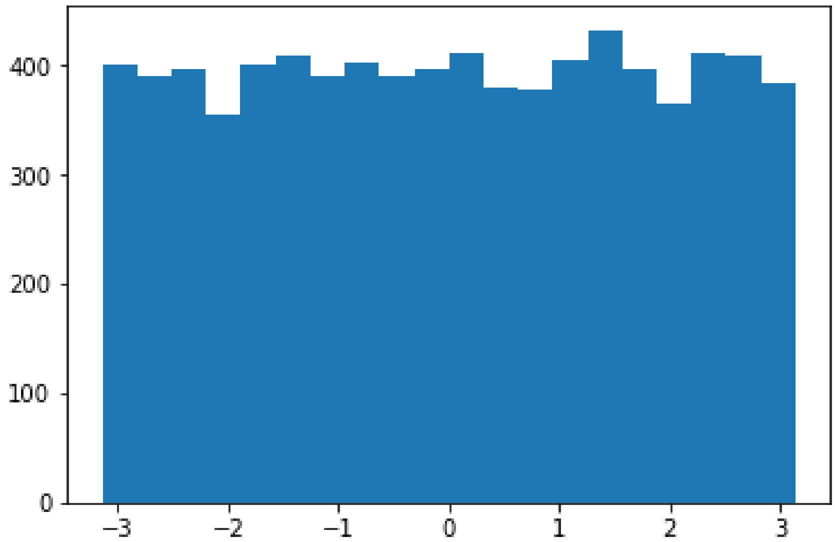
From Remark 10.12 we see that if we could sample from the unit disk, then the projection trick will produce samples uniform on the unit circle. How do we sample from the unit disk? We can use the concept of rejection sampling Algorithm 1, i.e. our sampling density is $\text{Uniform}([-1,1]^d)$ and our target density is $\frac{1}{|B_1|}\mathbb{1}_{B_1}(x)$.

Exercise 10.14. *The rejection sampling suggested above, is that equivalent to sampling from* $\text{Uniform}([-1,1]^d)$ *and accepting only those samples that lie inside the unit disk?*

[281]:
```
XY = np.random.uniform(-1,1,size=(10000,2))
XY_inCircle = XY[np.linalg.norm(XY,axis=1) < 1]
```

```
XY_inCircle = XY_inCircle / np.linalg.
↪norm(XY_inCircle,axis=1).reshape(-1,1)

import pylab
_=pylab.hist(np.arctan2(XY_inCircle[:,1],XY_inCircle[:
↪,0]),bins=20)
```



We know from Remark 10.12 that this is uniform on the unit circle, however is it a reasonable approach in higher dimension? We already showed that the volume of the unit ball decreases rapidly with dimension while the volume of the cube is 2^d, so the probability of being inside the unit ball is decreasing very rapidly,

Exercise 10.15. *What happens with the rejection sampling algorithm above when d is large?*

10.4.2 Uniform at random on the unit sphere in high dimension

What was the problem that we had, well basically if we sample from the unit square that distribution is not rotationally symmetric, thus if we sample from a rotationally symmetric distribution then it does not matter, we can just scale any sample to be on the unit circle. A prime example of a rotationally symmetric random variable is the multidimensional Gaussian.

Lemma 10.16. *Let Z be a d dimensional spherical Gaussian (see Model 10.3) then the density for f is rotationally symmetric (see Definition 10.13), as such the density for $\pi(Z)$ is uniform on the unit sphere S_1, where π is*

$$\pi(z) = \frac{z}{|z|}.$$

Exercise 10.17. *Prove the above lemma using radial coordinates.*

10.4.3 Uniform at random from the unit ball B_1?

In the above we learned how to generate uniform at random from the unit sphere. How can we use this to fill out the entire ball? Perhaps we think that if we take $r \sim \text{Uniform}([0,1])$ and $\theta \sim \text{Uniform}(S_1)$, do we then get $r\theta \sim \text{Uniform}(B_1)$? We know that $r\theta \mid r$ is uniform on S_r but $r\theta$ is not uniform, it turns out that we need to scale r as seen in the following theorem.

Theorem 10.18. *Let the dimension $d > 1$ be fixed, let $r \sim \text{Uniform}([0,1])$ and let $\theta \sim \text{Uniform}(S_1)$. Then*

$$r^{\frac{1}{d}}\theta \sim \text{Uniform}(B_1).$$

Proof. Assume that $X \sim \text{Uniform}(B_1)$, represent X in polar coordinates, i.e. $r_X\theta_X$, where $r_X \in [0,1]$ and $\theta_X \in S_1$. We know that given r_X the distribution for θ_X is uniform on the unit sphere. Secondly we know that r_X and θ_X are independent. The goal is to compute the density of r_X:

$$F_{r_X}(r) = \mathbb{P}(r_X \leq r) = \mathbb{P}(r_X\theta_X \in B_r) = \frac{|B_r|}{|B_1|} = \frac{r^d|B_1|}{|B_1|} = r^d,$$

where we just used the definition of the uniform distribution Model 10.11 and Lemma 10.7. If we now wish to sample from F_{r_X} we can use the inversion sampling technique (Theorem 5.38) and note that if $r \sim \text{Uniform}([0,1])$, then $F_{r_X}^{-1}(r) \sim F_{r_X}$. This proves our theorem. □

Remark 10.19. *The scaling of $r^{1/d}$ where $r \sim \text{Uniform}([0,1])$ tells us that we are more likely to get points with radius close to 1, than we are getting points with radius close to 0. This points to the fact that the most interesting things happen close to the unit sphere.*

10.5 High dimensional annulus theorem

The interesting thing is that in high dimension, random variables tend to concentrate on a spherical shell. Remember that for a d-dimensional spherical Gaussian X with standard deviation 1 in each dimension satisfies

$$\mathbb{E}\left[|X|^2\right] = d.$$

Thus one could expect that $|X|$ concentrates around $\sqrt{d}$, as hinted to in Lemma 10.5. The next theorem makes this sharper.

Theorem 10.20. *For a d-dimensional RV X with mean 0, with each component, sub-Gaussian with parameter 1 and variance a^2, then for any $\beta \leq \sqrt{d}$ we have*

$$\mathbb{P}\left(\sqrt{d}|a| - \beta \leq |X| \leq \sqrt{d}|a| + \beta\right) < 2e^{-\frac{\beta^2}{128}}.$$

Proof. First note that $|X|^2 = \sum_{i=1}^{d} |X_i|^2$ is the sum of independent random variables, and since X_i is sub-Gaussian with parameter 1, $|X_i|^2$ is sub-exponential with parameter 8, Lemma 3.15. Hence a simple application of Theorem 3.14 tells us that

$$\mathbb{P}\left(\frac{1}{d}|X|^2 - \frac{1}{d}\mathbb{E}\left[|X|^2\right] > \epsilon\right) < e^{-\frac{\epsilon^2 d}{128}} \wedge e^{-\frac{(\epsilon+1)d}{16}}$$

however

$$\mathbb{P}(|X|^2 > a^2 d + d\epsilon) > \mathbb{P}(|X| > |a|\sqrt{d} + \sqrt{d\epsilon})$$

so, denoting $\beta = \sqrt{d\epsilon}$ and since

$$\frac{(\epsilon+1)d}{16} = \frac{(\beta^2/d+1)d}{16} = \frac{\beta^2+d}{16} > \frac{\beta^2}{16}$$

and

$$\frac{\epsilon^2}{128} = \frac{\beta^4/d}{128} < \frac{\beta^4/\beta^2}{128} = \frac{\beta^2}{128} < \frac{\beta^2}{16}$$

hence we get

$$P(|X| > |a|\sqrt{d} + \beta) < e^{-\frac{\beta^2}{128}}.$$

The other side of the inequality is obtained in a similar way, and, together with the union bound gives the result. □

10.6 Bibliography

This section is losely built on [BlHo].