STAT 672: Homework #1

Tom Wallace

February 10, 2018

Problem 1

A

See attached code `hw1_A.py` for an implementation of rejection sampling from the unit cube.

The acceptance probability is the volume of an n-dimensional ball divided by the volume of an n-dimensional cube:

$$P_{X\sim B^d_\infty}(||X||_2 \leq 1) = \frac{\mathrm{Vol}(B_2^d)}{\mathrm{Vol}(B_\infty^2)} = \frac{\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}}{2^d} \tag{1}$$

We know that $\Gamma(n+1) = n!$ Thus, $\Gamma(\frac{d}{2}+1) = (\frac{d}{2})!$

Stirling's approximation for factorials is useful here.

$$n! \approx \sqrt{2\pi n} n^n e^{-n} \tag{2}$$

Let us consider the case of $2d$. In this case:

$$\mathrm{Vol}(B_2^{2d}) = \frac{\pi^d}{\Gamma(d+1)} = \frac{\pi^d}{d!} \approx \frac{\pi^d}{\sqrt{2\pi d} d^d e^{-d}} \tag{3}$$

Rearranging terms, we have:

$$= \frac{1}{\sqrt{2\pi d}} \left(\frac{\pi e}{d}\right)^d \tag{4}$$

Referring to $\left(\frac{\pi e}{d}\right)^d$, the denominator grows (much) faster with d than does the numerator. As a consequence, if we extend d infinitely, the limit of the ratio is 0.

$$\lim_{d\to\infty} \left(\frac{\pi e}{d}\right)^d = 0 \tag{5}$$

Which implies:

$$\lim_{d\to\infty} \frac{1}{\sqrt{2\pi d}} \left(\frac{\pi e}{d}\right)^d = \frac{1}{\infty} \times 0 = 0 \tag{6}$$

Returning to (1), we compute the the volume of the unit cube as $d \to \infty$.

$$\lim_{d\to\infty} 2^{2d} = \infty \tag{7}$$

So, the ratio of the unit sphere to the unit cube is:

$$\lim_{d\to\infty} \frac{\mathrm{Vol}(B_2^d)}{\mathrm{Vol}(B_\infty^2)} = \frac{0}{\infty} = 0 \tag{8}$$

Since this ratio also is the probability of accepting a sample in a rejection sampling scheme, our expected runtime to generate a single random vector from B_2^d with rejection sampling grows asymptotically with d. For example, for $d = 10000$, our program would run for a very, very, *very* long time.

B

See attached code `hw1_1B.py` for an implementation of polar sampling.

C

The methods implemented in **A** and **B** can be empirically compared. Each program was used to generate 100 samples with $d = 10$. The execution runtime was measured with the Linux `time` command, e.g. `time python3 hw1_1A.py`. Method B is significantly faster than Method A, as shown in Table 1.

Table 1: Runtime comparison (seconds)

	A	**B**
real	2.131	0.744
user	2.150	0.744
sys	0.172	0.192

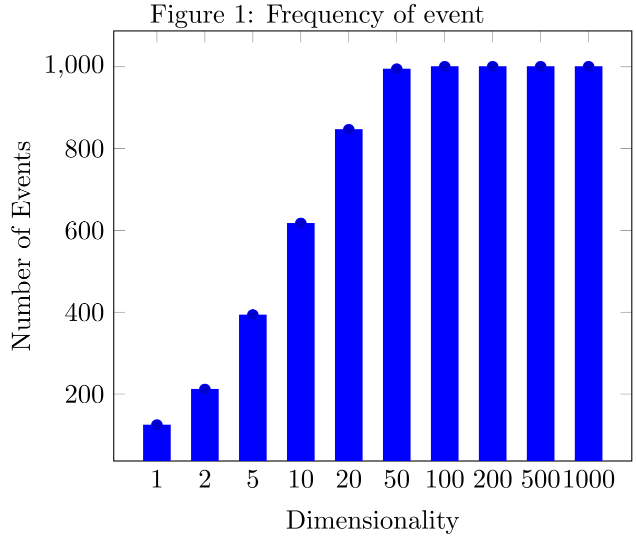
Problem 2

A

Simulation code is attached as `hw1_2A.py`.

As d grows larger, the frequency of the event $\{||X_+ - Z||_2^2 \geq ||X_- - Z||_2^2\}$ increases. This is somewhat counter-intuitive. Because Z is drawn from the same distribution as X_+, we would expect their difference $X_+ - Z$ to result in something close to a zero-vector, which should have a smaller Euclidean norm than that of $X_- - Z$ (since these two have different means), resulting in *low* frequency of the event. This is true in low dimensions, but as dimensionality grows, the event occurs with *high* frequency. Why this counter-intuitive result occurs is explained in **B**.

Figure 1: Frequency of event



B

$$E[||X_+ - Z||_2^2]$$

$$= E[||X_+||_2^2] - 2E[X_+^T Z] + E[||Z||_2^2]$$
$$E[||X_+||_2^2] = ||\mu_+||_2^2 + tr(\Sigma_+) = 25 + 4d$$
$$E[||Z||_2^2] = E[||X_+||_2^2] = 25 + 4d$$
$$E[X_+^T Z] = E[X_+^T]E[Z] = ||\mu_+||^2 = 25$$
$$E[||X_+ - Z||_2^2] = 2(25 + 4d) - 2(25) = \boxed{8d}$$

$$E[||X_- - Z||_2^2]$$
$$= E[||X_-||_2^2] - 2E[X_-^T Z] + E[||Z||_2^2]$$
$$E[||X_-||_2^2] = ||\mu_-||_2^2 + tr(\Sigma_-) = 25 + d$$
$$E[||Z||_2^2] = 25 + 4d$$
$$E[X_-^T Z] = E[X_-^T]E[Z] = -25$$
$$E[||X_- - Z||_2^2] = (25 + d) - 2(-25) + (25 + 4d) = \boxed{100 + 5d}$$

This theoretical result explains our empirical observations in **A**. With low dimensionality, the expected value of $||X_- - Z||_2^2$ is greater than that of $||X_+ - Z||_2^2$, and so our event occurs with low frequency. However, the expected value of $||X_+ - Z||_2^2$ grows faster with d, and so in high dimensions, our event occurs with high frequency. In essence, concentration of measure / curse of dimensionality magnify the effect of X_+'s higher variance as dimensionality grows higher. The different rate of growth of expected value with d is visualized below in Figure 2.

Figure 2: Change in expected value with d