Problem Set 1

Javier Mas Adell

1.

We need to prove that $|M - m| \leq \sqrt{2}\sigma$. We know that, by Chebyshev's inequality and using strict inequalities:

$$P(|x - m| > \sqrt{2}\sigma) < \frac{\sigma^2}{(\sqrt{2}\sigma)^2}$$

Which gives us the interval delimited by

$$P(x < m - \sqrt{2}\sigma) < \frac{1}{2}$$
$$P(x > m + \sqrt{2}\sigma) < \frac{1}{2}$$

Therefore,

$$P[x \notin (m - \sqrt{2}\sigma, m + \sqrt{2}\sigma)] < \frac{1}{2}$$

and, its complementary:

$$P[x \in (m - \sqrt{2}\sigma, m + \sqrt{2}\sigma)] \geq \frac{1}{2}$$

We know that, by its definition, the median will lie somewhere within that interval and, therefore, cannot be further than $\sqrt{2}\sigma$ away from the mean. Hence, $|M - m| \leq \sqrt{2}\sigma$.

2.

The ex2.R file includes a function that takes arguments N (sample size), d (precision) and the number of samples and returns a table like the ones attached at the end of the document. Through this code, I test the performance of the MoM estimator, compared to the sample mean, in terms of average deviation and worst case deviation in standard-like distributions but also in heavy-tailed and skewed distributions. More specifically, I assess their relative performance in: (i) the normal distribution, (ii) the student's t-distribution, (iii) the pareto distribution and (iv) the weibull distribution. All of them with varying parameters and sample sizes.

There are four key ideas we can extract from these tables:

1. The mean performs better than the MoM in most cases. The average deviation of the mean from the expected value turns out to be lower in most scenarios.

2. However, the difference in average deviation between mean and MoM is extremely low. Both tend to perform well in those scenarios where the mean is the winner in terms of performance.

3. In highly (highly) skewed and heavy-tailed distributions, such as the normal distribution with $\sigma^2 = 25$, the student's t-distribution with $df = 1$ or the pareto with $\alpha = 2$, the MoM performs substantially better than the mean. Actually, in these distributions, the mean shows worst case scenarios with extreme deviations from the actual expected value. However, even in skewed but not-so-skewed distributions like Weibull with high parameters, the mean performs well as an estimator of the expected value.

4. As we could expect, the mean performs better as we increase the sample size. The outperformance of the MoM disappears in some of the distributions above mentioned. However, the mean is still a dangerous estimator in heavy-tailed distributions like student's t with 1 df, with a worst-case deviation of 1409.

These results are reaffirmed when we try with other distributions, such as the poisson distribution, whose results I have not shown to keep tables shorter. As a conclusion, we can see that in normal cases the mean outperforms than the MoM, but not by a substantial quantity. The mean, however, turns out to be a risky estimator in heavy-tailed or highly-skewed distributions, specially with small sample sizes. As we have discussed in class, it might be recommendable to use the MoM estimator when we cannot make the necessary assumptions for the sample mean to work.

3.

We want to prove that

$$P\left\{\frac{1}{n}\sum_{i=1}^{n} X_i < m - t\right\} \leq e^{-\frac{nt^2}{2a^2}}$$

If we play a little bit with the left hand side we can reach this

$$P\left\{\sum_{i=1}^{n} X_i - mn < -tn\right\}$$

We are given that $X_1, \ldots, X_n$ are i.i.d. random variables with $E[X_1] = m = E[X_i]$. Therefore, $nm = \sum_{i=1}^{n} m$. With this beautiful conclusion, we reach an equation that can work with Chernoff bounds in the way we want

$$\begin{aligned}
P\left\{\sum_{i=1}^{n} E[X_i] - \sum_{i=1}^{n} X_i > tn\right\} &\leq \frac{E[\prod_{i=1}^{n} e^{\lambda(E[X_i]-X_i)}]}{e^{\lambda tn}} \\
&= \frac{1}{e^{\lambda tn}} \prod_{i=1}^{n} E[e^{\lambda(E[X_i]-X_i)}] \\
&= \frac{e^{\lambda mn}}{e^{\lambda tn}} \prod_{i=1}^{n} E[e^{-\lambda X_i}] \\
&= e^{\lambda n(m-t)} \prod_{i=1}^{n} E[e^{-\lambda X_i}]
\end{aligned}$$

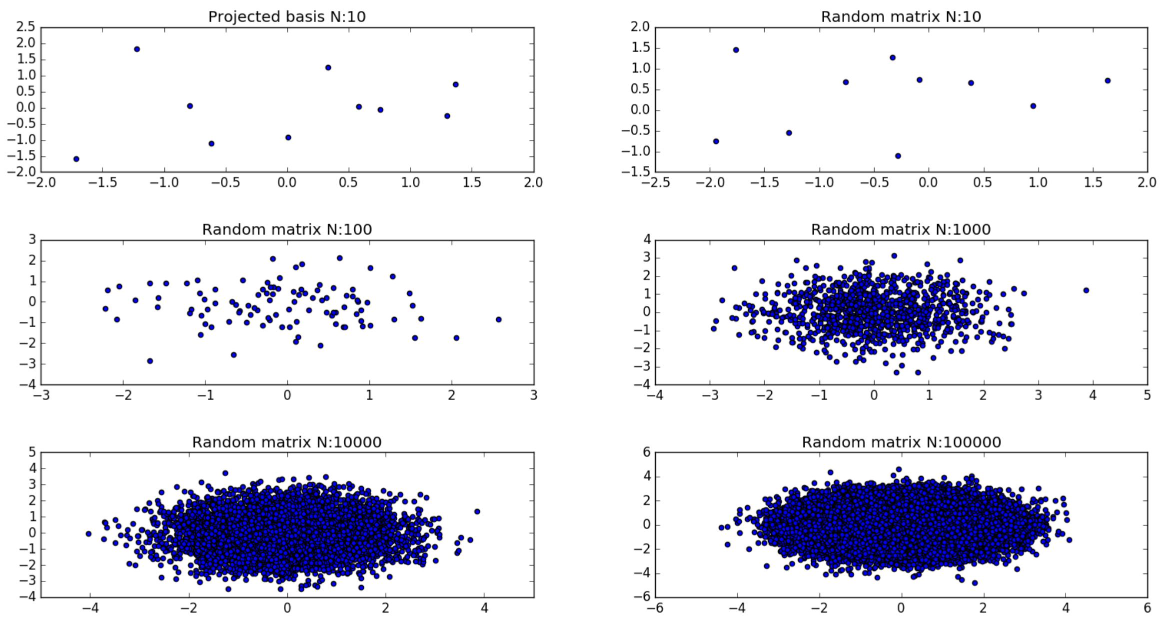
Hm... it seems we are stuck... wait, we are given that $e^{-x} \leq 1 - x + \frac{x^2}{2} \forall x > 0$ and our X_i are non-negative... Nice, let us see where this takes us.

$$e^{\lambda n(m-t)} \prod_{i=1}^{n} E[e^{-\lambda X_i}] \leq \prod_{i=1}^{n} E[e^{1-\lambda X_i + \frac{\lambda X^2}{2}}]$$
$$= \prod_{i=1}^{n} 1 - \lambda E[X] + \frac{\lambda^2}{2} E[X^2]$$

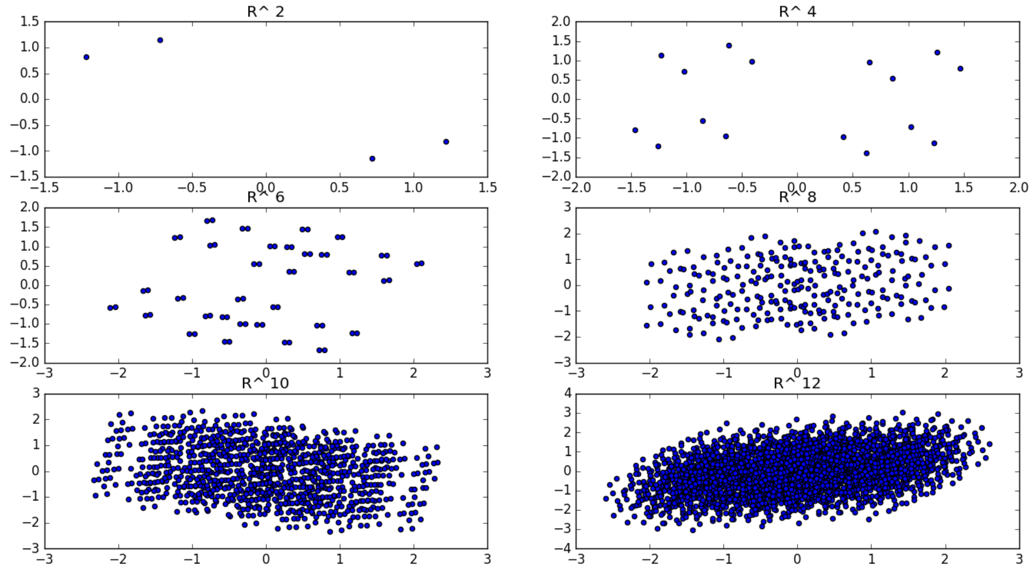
4.

The file ex4.py includes the code needed to complete the tasks and produce the plots for this exercise.

The following group of plots show the projection of 10 basis vectors along with several sets of random draws. Given that a matrix of basis vectors in a specific order is, simply, the identity matrix, the resulting projected matrix is merely the plot of 20 random draws from a normal distribution with $\sigma^2 = 1/d$. From this plot it is difficult to see whether pairwise distances have been preserved.



For the second part of the exercise, the following plot projects the vertices of a hypercube in d-dimensions to 2-dimensions.



The result, as we can see, is a sort of projection of the shape of the hypercube to the $d = 2$. Among different possibilities, I have displayed the clearest case, the square, where we can see something similar to a square that has been twisted into some kind of parallelogram due to the projection. It is the clearest case because we can easily compare the values and check that, indeed, the pairwise distances have been magically preserved. The vertices of a unit polyhedron of two dimensions are the four possible combinations of $\{-1, 1\}$. That is, for a vertex in $\{-1, 1\}$ we get also its opposite $\{1, -1\}$. Similarly, here, we can see that, for a data point $\sim \{0.8, -1.25\}$, we get also its opposite $\sim \{-0.8, 1.25\}$. As D goes higher, it gets more difficult to observe this fact from the plot, but we can still see how the vertices group in sets of two and are, overall, vertically and horizontally symmetric (and also closer as the scale of the plot grows with the number of points).

Table 1: N= 20, D=0.1, samples=1000

	Distribution	Varying parameter	Exp Value	MoM	Mean	Avg MoM dev	Avg mean dev	Worst case MoM	Worst case mean
1	Normal	Variance = 1	0.00	-0.01	-0.00	0.22	0.18	0.86	0.67
2	Normal	Variance = 4	0.00	0.03	0.00	0.87	0.69	3.33	2.51
3	Normal	Variance = 9	0.00	-0.04	-0.09	1.93	1.55	6.94	6.54
4	Normal	Variance = 16	0.00	0.06	0.00	3.31	2.68	17.70	12.02
5	Normal	Variance = 25	0.00	0.57	0.31	5.51	4.44	22.47	16.85
6	T-Student	Df = 1	0.00	0.00	-0.07	0.31	5.61	1.45	922.52
7	T-Student	Df = 4	0.00	0.01	0.01	0.25	0.25	1.08	1.33
8	T-Student	Df = 9	0.00	0.01	0.00	0.22	0.19	0.85	0.87
9	T-Student	Df = 16	0.00	0.02	0.01	0.23	0.20	0.91	0.79
10	T-Student	Df = 25	0.00	-0.01	-0.01	0.23	0.18	0.91	0.76
11	Pareto	Alpha = 2	2.00	0.47	1.00	1.53	1.09	1.92	4.28
12	Pareto	Alpha = 5	1.25	0.64	1.00	0.61	0.32	1.06	1.24
13	Pareto	Alpha = 10	1.11	0.70	1.00	0.43	0.22	0.89	0.86
14	Pareto	Alpha = 17	1.06	0.71	1.00	0.37	0.19	0.88	0.90
15	Pareto	Alpha = 26	1.04	0.72	1.00	0.34	0.18	0.90	0.97
16	Weibull	K = 25	0.98	0.98	0.98	0.01	0.01	0.04	0.04
17	Weibull	K = 36	0.98	0.99	0.98	0.01	0.01	0.03	0.03
18	Weibull	K = 49	0.99	0.99	0.99	0.01	0.00	0.02	0.02
19	Weibull	K = 64	0.99	0.99	0.99	0.00	0.00	0.02	0.02
20	Weibull	K = 81	0.99	0.99	0.99	0.00	0.00	0.02	0.01

Table 2: N=100, d=0.1, samples=1000

	Distribution	Varying parameter	Exp Value	MoM	Mean	Avg MoM dev	Avg mean dev	Worst case MoM	Worst case mean
1	Normal	Variance = 1	0.00	-0.00	-0.00	0.09	0.08	0.41	0.37
2	Normal	Variance = 4	0.00	0.00	0.01	0.39	0.31	1.73	1.41
3	Normal	Variance = 9	0.00	-0.04	-0.02	0.87	0.70	3.38	2.56
4	Normal	Variance = 16	0.00	-0.11	-0.04	1.52	1.28	7.41	5.53
5	Normal	Variance = 25	0.00	0.05	0.02	2.29	1.91	10.32	7.20
6	T-Student	Df = 1	0.00	0.01	0.82	0.31	5.90	1.79	1126.40
7	T-Student	Df = 4	0.00	0.00	0.00	0.12	0.11	0.51	0.48
8	T-Student	Df = 9	0.00	-0.00	-0.01	0.11	0.09	0.43	0.38
9	T-Student	Df = 16	0.00	-0.00	0.00	0.10	0.08	0.44	0.37
10	T-Student	Df = 25	0.00	-0.00	-0.00	0.10	0.08	0.44	0.30
11	Pareto	Alpha = 2	2.00	0.75	1.02	1.25	1.04	1.59	17.66
12	Pareto	Alpha = 5	1.25	0.90	1.00	0.35	0.25	0.73	0.69
13	Pareto	Alpha = 10	1.11	0.93	1.00	0.19	0.13	0.50	0.42
14	Pareto	Alpha = 17	1.06	0.93	1.00	0.15	0.10	0.46	0.37
15	Pareto	Alpha = 26	1.04	0.94	1.00	0.13	0.09	0.48	0.34
16	Weibull	K = 25	0.98	0.98	0.98	0.00	0.00	0.02	0.02
17	Weibull	K = 36	0.98	0.99	0.98	0.00	0.00	0.02	0.01
18	Weibull	K = 49	0.99	0.99	0.99	0.00	0.00	0.01	0.01
19	Weibull	K = 64	0.99	0.99	0.99	0.00	0.00	0.01	0.01
20	Weibull	K = 81	0.99	0.99	0.99	0.00	0.00	0.01	0.01

Table 3: N = 100000, d = 0.1, samples=1000

	Distribution	Varying parameter	Exp Value	MoM	Mean	Avg MoM dev	Avg mean dev	Worst case MoM	Worst case mean
1	Normal	Variance = 1	0.00	-0.00	0.00	0.00	0.00	0.01	0.01
2	Normal	Variance = 4	0.00	-0.00	-0.00	0.01	0.01	0.04	0.04
3	Normal	Variance = 9	0.00	-0.00	0.00	0.03	0.02	0.14	0.14
4	Normal	Variance = 16	0.00	-0.00	-0.00	0.05	0.04	0.19	0.15
5	Normal	Variance = 25	0.00	-0.00	-0.00	0.08	0.06	0.35	0.23
6	T-Student	Df = 1	0.00	0.00	-3.49	0.31	6.33	1.83	1409.84
7	T-Student	Df = 4	0.00	0.00	0.00	0.00	0.00	0.02	0.02
8	T-Student	Df = 9	0.00	-0.00	-0.00	0.00	0.00	0.02	0.01
9	T-Student	Df = 16	0.00	-0.00	-0.00	0.00	0.00	0.01	0.01
10	T-Student	Df = 25	0.00	0.00	0.00	0.00	0.00	0.01	0.01
11	Pareto	Alpha = 2	2.00	0.99	1.00	1.01	1.00	1.04	1.03
12	Pareto	Alpha = 5	1.25	1.00	1.00	0.25	0.25	0.27	0.26
13	Pareto	Alpha = 10	1.11	1.00	1.00	0.11	0.11	0.12	0.12
14	Pareto	Alpha = 17	1.06	1.00	1.00	0.06	0.06	0.08	0.07
15	Pareto	Alpha = 26	1.04	1.00	1.00	0.04	0.04	0.05	0.05
16	Weibull	K = 25	0.98	0.98	0.98	0.00	0.00	0.00	0.00
17	Weibull	K = 36	0.98	0.98	0.98	0.00	0.00	0.00	0.00
18	Weibull	K = 49	0.99	0.99	0.99	0.00	0.00	0.00	0.00
19	Weibull	K = 64	0.99	0.99	0.99	0.00	0.00	0.00	0.00
20	Weibull	K = 81	0.99	0.99	0.99	0.00	0.00	0.00	0.00