Name: Hongda Li
Class: CSE 546

B.1

Let's start by looking for the probability of observing $Y \le y$. Notice that if $Y \le y$, then it means $\forall 1 \le i \le n$, we have $x_i \le y$, where,l x_i is a observed sample for the rv X_i.

Then it means that:

$$\mathbb{P}(Y \le y) = y^n \tag{B.1.1}$$

This is true because, y set a threshold for what x_i could be. By the assumption that x is unif distributed in interval $[0,1]$, and the fact that all of them is less than y, we use the probability for independent event. The probability of observing $x_i \le y$ each time is y, and there are n such an event, hence y^n. Mathematically represented as:

$$\mathbb{P}(\max(X_1, X_2, \cdots, X_n) \le y) = \prod_{i=1}^{n} \mathbb{P}(X_i \le y) = y_n \tag{B.1.2}$$

Take note that y^n is the CDF, and the PDF will be the derivative of it, giving us: ny^{n-1}. And it's not hard to verify that it makes sense for $n = 1$, giving using the uniform distribution on $[0,1]$.

Then the expected value for random variable Y will be given by:

$$\begin{aligned}\mathbb{E}[Y] &= \int_0^1 yny^{n-1}dy \qquad \text{(B.1.3)}\\ &= \int_0^1 ny^n dy \\ &= \frac{n}{n+1}\end{aligned}$$

B.2

Note that, in order to use the markov inequality, we will have to make the random variable non negative. We were given that $x > 0$, so we can say that:

$$\begin{aligned}\mathbb{P}(X \ge \mu + \sigma x) =& \mathbb{P}\left(\frac{x-\mu}{\sigma} \ge x\right) \qquad \text{(B.2.1)}\\ \mathbb{P}\left(\left(\frac{X-\mu}{\sigma}\right)^2 \ge x^2\right) \le& \frac{\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]}{x^2}\end{aligned}$$

Here, we are just exchanging the variables and applying the Markov Inequality. In addition, notice that $\mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$, which is just the variance.

Then we have:

$$\mathbb{E}\left[\left(\frac{X-\mu}{\sigma^2}\right)\right] = \frac{1}{\sigma^2}\mathbb{E}\left[(X-\mu)^2\right] = \frac{\sigma^2}{\sigma^2} = 1 \tag{B.2.2}$$

Substituting results in B.2.2 into B.2.1, we have the desired expression which is:

$$\mathbb{P}(X \le \mu + \sigma x) \le \frac{1}{x^2} \tag{B.2.3}$$

B.3

The trace of a matrix is the sum of all its diagonal elements, the trace of a matrix X is denoted as $\text{Tr}[X]$. Here, we define that A as an $n \times m$ matrix and B as a $m \times n$ matrix and we want to show that $\text{Tr}[AB] = \text{TR}[BA]$, which is just:

$$\begin{aligned}
\text{Tr}[AB] =& \sum_{k=1}^{n}(AB)_{k,k} \\
=& \sum_{k=1}^{n}\sum_{i=1}^{m} A_{k,i}B_{i,k} \\
=& \sum_{i=1}^{m}\sum_{k=1}^{n} B_{i,k}A_{k,i} \\
=& \sum_{i=1}^{m}(BA)_{i,i} \\
=& \text{Tr}[BA]
\end{aligned} \tag{B.3.1}$$

B.4

Consider a bunch of non zero vector $[v_1, v_2, \cdots v_n]$ and these column vectors are horizontally stacked into matrix V. Each of the vector v_i is in $\mathbb{R}^d s$

(B.4.a)

What is the minimum and maximal rank of $\sum_{i=1}^{n} v_i v_i^T$?

For any none-zero vector v_i, the outter product wih itself will be a rank-1 matrix. This is the case because, given any vector b, $b \in \mathbb{R}^d$, the outter product with itself is:

$$bb^T = \begin{bmatrix} b_1 b & b_2 b & \cdots & b_d b \end{bmatrix} \tag{B.4.a.1}$$

The above matrix is rank-1, because the columns are multiple with each other.

It's not hard to see that, if we sum up outter products of vector with itself, we are just taking some kind of linear combinations of each of the vector v_i on each column of the matrix: $\sum_{i=n}^{m} v_i v^T$. Nothing can demonstrate this better than writing it down, here we denote $(v_i)_j$ to be the j th element of the vector v_i, which will be:

$$\sum_{i=1}^{n} v_i v^T = \begin{bmatrix} \sum_{i=1}^{n}(v_i)_1 v_i & \sum_{i=1}^{n}(v_i)_2 v_i & \cdots & \sum_{i=1}^{n}(v_i)_n v_i \end{bmatrix} \tag{B.4.a.2}$$

The maximum number of rank possible for $\sum_{i=1}^{n} v_i v^T$ is $\min(d, n)$. If $n > d$, then these rank-1 matrix can span the whole $\mathbb{R}^d$.

Let's consider the case when $n \leq d$, then let $v_i = e_i$, where e_i is the basis standard basis vector. And it's not hard to show that $e_i e_i^T - e_i$, therefore, the matrix $\sum_{i=1}^{n} e_i e_i^T$ has ones on is diagonal. Such a matrix has a rank of n.

The minimum rank for the matrix $\sum_{i=1}^{n} v_i v^T$ is 1. Because all vector are non zeros, and it happens when all the vector v_i is the same for $1 \leq i \leq n$.

(B.4.b)

The maximum rank of the matrix V is $\min(n, d)$. The maximal span is made by considering matrix V linear Independence columns.

The minimum rank for the matrix V is 1, by the fact that all columns are non-zero, and the worst case is when all the vectors are all the same, giving us a rank of 1 for the matrix V.

(B.4.c)

Let's denote S to be the matrix $\sum_{i=1}^{n}(Av_i)(Av_i)^T$.

Each column of the matrix S is linear combinations of vectors Av_i for $1 \leq i \leq n$.

Notice that when the matrix A is full rank, then it's column space has a dimension of d. because the matrix is skinny.

To demonstrate the maximal span for the matrix S, let's consider the case where A has zeros down its diagonal. Since $D > d$, we will have the last $D - d$ to D rows full of zeros.

Under that case, the vector Av_i becomes $v_i \in \mathbb{R}^d$ but the $D - d$ to D entries are all zeros.

Therefore, we have the exact same argument in part (b), we just let $v_i = e_i$, then the matrix S will have a maximal span of $\min(d, n)$.

The minimum rank for matrix S is 1, since all vector can not be zero, then we can choose all v_i to be the same vector, then matrix S has a rank of one.

(B.4.d)

The maximum rank for the matrix AV is achieved when the matrix V is full rank, having a rank of d, then, and the matrix A is full-rank as well. Then the rank of matrix AV is d.

This is the case by considering any vector y that is spanned by the columns of matrix A, for each such y, is some vector $x \in \mathbb{R}^d$ such that $Ax = y$, and x will be unique. Then, since V span $\mathbb{R}^d$, then there exists a linear combinations of columns of V that equals to x (say the L.C is vector u). Therefore, the system $AVu = y$ has a solution vector u to it.