Homework 0 A

Fall 2020, CSE 546: Machine Learning
John Franklin Crenshaw
October 5, 2020

Probability and Statistics

A.1 (Bayes Rule, from Murphy exercise 2.4) After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease? (Show your calculations as well as giving the final result)

Let A denote having the disease, and B denote testing positive for the disease. We are given $p(B|A) = 0.99$, $p(A) = 10^{-4}$. Using these, we can calculate

$$p(B) = p(B|A)p(A) + p(B|\neg A)p(\neg A) = 0.99 \cdot 10^{-4} + (1 - 0.99) \cdot (1 - 10^{-4}) \approx 0.01$$

Finally, using Bayes' Theorem, we have.

$$p(A|B) = p(B|A)\ \frac{p(A)}{p(B)} = 0.99 \cdot \frac{10^{-4}}{0.01} \approx 0.01$$

So there's only a 1% chance you actually have the disease.

A.2 For any two random variables X, Y the *covariance* is defined as $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[x])(Y - \mathbb{E}[Y])]$.

a. If $\mathbb{E}[Y|X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

$$\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[XY - X\,\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\,\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\
\mathbb{E}[XY] &= \int xy\,p(x, y)\,dx\,dy = \int xy\,p(y|x)\,p(x)\,dx\,dy \\
&= \int x\,p(x)\left[\int y\,p(y|x)\,dy\right]dx = \int x^2\,p(x)dx = \mathbb{E}[X^2] \\
\mathbb{E}[Y] &= \int y\,p(y)\,dy = \int y\,p(y|x)\,p(x)\,dy\,dx \\
&= \int p(x)\left[\int y\,p(y|x)\,dy\right]dx = \int x\,p(x)\,dx = \mathbb{E}[X]
\end{aligned}$$

where I substituted $\mathbb{E}[Y|X = x] = x$ for the integrals in brackets above. Finally,

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\,\mathbb{E}[X] + \mathbb{E}[X]^2] \\
&= \mathbb{E}[X^2] - 2\,\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \text{Cov}(X, Y),
\end{aligned}$$

which proves the desired result.

b. If X, Y are independent, show that $\text{Cov}(X, Y) = 0$.

Above I showed $\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. If X, Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, as the integrals factor. Thus $\text{Cov}(X,Y) = \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0$

A.3 Let X and Y be independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$.

a. Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$

We can write $h(z)$ as an integral over values of x:

$$h(z) = \int_{-\infty}^{\infty} p(z|x)f(x)dx.$$

Now as $Z = X + Y$, the conditional probability $p(z|x)$ is the same as asking what's the probability that $y = z - x$, i.e. $p(z|x) = g(z-x)$. Putting it all together,

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx.$$

b. If X and Y are both independent and uniformly distributed on $[0,1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0,1]$ and 0 otherwise), what is h, the PDF of $Z = X + Y$?

Clearly $h(z) = 0$ when $z < 0$ or $z > 2$, as you are adding two random numbers between 0 and 1. Now for $g(z-x) \neq 0$, we must have $z - 1 \leq x \leq z$. Combining these limits with the limits $0 \leq x \leq 1$, we have two different cases:

$$z < 1: \quad h(z|z<1) = \int_0^z dz = z$$
$$z \geq 1: \quad h(z|z\geq 1) = \int_{z-1}^1 dz = 2 - z$$

So the PDF for z is

$$h(z) = \begin{cases} z & 0 < z \leq 1 \\ 2 - z & 1 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

A.4 A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2. Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0,1)$.

$$\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b = a\mu + b = 0$$
$$\text{Var}[Y] = \text{Var}[aX + b] = a^2\,\text{Var}[X] = a^2\sigma^2 = 1$$
$$\Rightarrow \quad a = \sigma^{-1}, \quad b = \frac{\mu}{\sigma}$$

A.5 For a random variable Z, its mean and variance are defined as $\mathbb{E}[Z]$ and $\mathbb{E}[(Z - \mathbb{E}[Z])^2]$, respectively. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, each with mean μ and variance σ^2. If we define $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$, what is the mean and variance of $\sqrt{n}(\hat{\mu}_n - \mu)$?

$$\mathbb{E}\left[\sqrt{n}(\hat{\mu}_n - \mu)\right] = \sqrt{n}\left(\mathbb{E}[\hat{\mu}_n] - \mu\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] - \mu\right)$$
$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mu - \mu\right) = \sqrt{n}(\mu - \mu) = 0$$
$$\text{Var}\left[\sqrt{n}(\hat{\mu}_n - \mu)\right] = \text{Var}\left[\sqrt{n}\hat{\mu}_n\right] = n\,\text{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i,j=1}^n \text{Cov}(X_i, X_j)$$
$$= \frac{1}{n}\sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n}\cdot n\sigma^2 = \sigma^2$$

In the last line I used the fact that the X_i are uncorrelated (i.e. $\text{Cov}(X_i, X_j) = 0,\ \forall i \neq j$).

A.6 If $f(x)$ is a PDF, the cumulative distribution function (CDF) is defined as $F(X) = \int_{-\infty}^{x} f(y)dy$. For any function $g : \mathbb{R} \to \mathbb{R}$ and random variable X with PDF $f(x)$, recall that the expected value of $g(x)$ is defined as $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(y)f(y)dy$. For a boolean event A, define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Thus, $\mathbf{1}\{x \leq a\}$ is 1 whenever $x \leq a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]$. Let $X_1, \ldots, X_n$ be *independent and identically distributed* random variables with CDF $F(x)$. Define $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}$. Note, for every x, that $\hat{F}_n(x)$ is an *empirical estimate* of $F(x)$. You may use your answers to the previous problem.

a. For any x, what is $\mathbb{E}[\hat{F}_n(x)]$?

$$\mathbb{E}[\hat{F}_n(x)] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq x\}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\mathbf{1}\{X_i \leq x\}] = \frac{1}{n}\sum_{i=1}^{n} F(x) = \frac{1}{n} \cdot n\, F(x) = F(x)$$

b. For any x, the variance of $\hat{F}_n(x)$ is $\mathbb{E}[(\hat{F}_n(x) - F(x))^2]$. Show that $\mathrm{Var}(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.

c. Using your answer to b, show that for all $x \in \mathbb{R}$, we have $\mathbb{E}[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$.

To establish the upper bound, we need to find the max of $F(x)(1 - F(x))$:

$$\frac{d}{dF(x)}(F(x)(1 - F(x))) = 1 - 2F(x) \longrightarrow \text{max when } F(x) = \frac{1}{2}.$$

Plugging this in, we get the desired result: $\mathbb{E}[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$.

Linear Algebra and Vector Calculus

A.7 (Rank) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. For each matrix A and B,

a. what is its rank?

The characteristic polynomial of A is $\det(A - \lambda I) = -\lambda^3 + 3\lambda^2 + 4\lambda = -\lambda(\lambda - 4)(\lambda + 1)$. As the characteristic polynomial has three distinct roots, A is diagonalizable. Since two of the roots are non-zero, $\mathrm{rank}(A) = 2$.

The characteristic polynomial of B is $\det(B - \lambda I) = -\lambda^3 + 3\lambda^2 + 4\lambda$, which is the same as A. Therefore, $\mathrm{rank}(B) = \mathrm{rank}(A) = 2$

b. what is a (minimal size) basis for its column span?

The dimension of the column space of a matrix is equal to its rank. Thus for each matrix, we need two basis vectors. We can see that the first two columns of both matrices are linearly independent of one another (and are the same for both matrices). So a basis for the column span of both matrices is

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

A.8 (Linear equations) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

a. What is Ac?

$$Ac = \begin{bmatrix} 0 \cdot 1 + 2 \cdot 1 + 4 \cdot 1 \\ 2 \cdot 1 + 4 \cdot 1 + 2 \cdot 1 \\ 3 \cdot 1 + 3 \cdot 1 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

b. What is the solution to the linear system $Ax = b$?

Using Gaussian elimination:

$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array}\right] \begin{array}{l} \cdot \frac{1}{2} \\ -2R_1 \\ -\frac{3}{2}R_1 \end{array} \rightarrow \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 2 & 0 & -6 & 2 \\ 3 & 0 & -5 & -1 \end{array}\right] \begin{array}{l} \cdot \frac{1}{2} \\ -\frac{3}{2}R_2 \end{array}$$

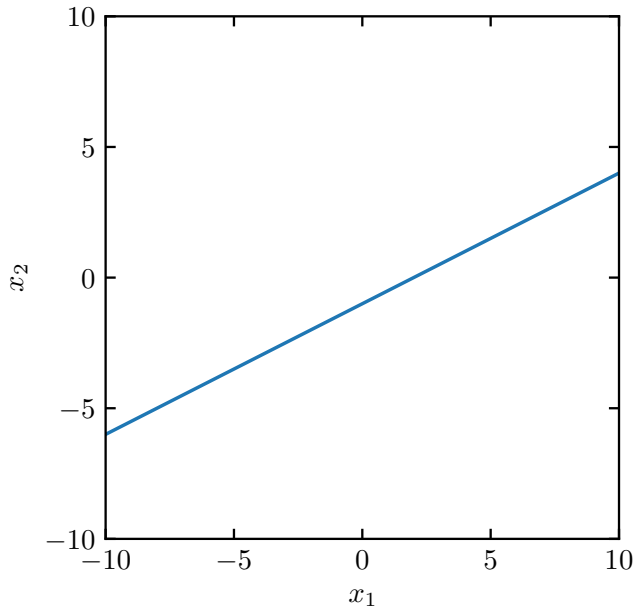
$$\rightarrow \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 0 & -3 & 1 \\ 0 & 0 & 4 & -4 \end{array}\right] \begin{array}{l} -\frac{1}{2}R_3 \\ +\frac{3}{4}R_3 \\ \cdot \frac{1}{4} \end{array} \rightarrow \left[\begin{array}{ccc|c} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

So $x = \begin{bmatrix} -2 & 1 & -1 \end{bmatrix}^T$.

A.9 (Hyperplanes) Assume w is an n-dimensional vector and b is a scalar. A hyperplane in $\mathbb{R}^n$ is the set $\{x : x \in \mathbb{R}^n, \text{ s.t. } w^T x + b = 0\}$.

a. ($n = 2$ example) Draw the hyperplane for $w = \begin{bmatrix} -1 & 2 \end{bmatrix}^T$, $b = 2$.

This is the equation $-x_1 + 2x_2 + 2 = 0$, or $x_2 = \frac{1}{2}x_1 - 1$.



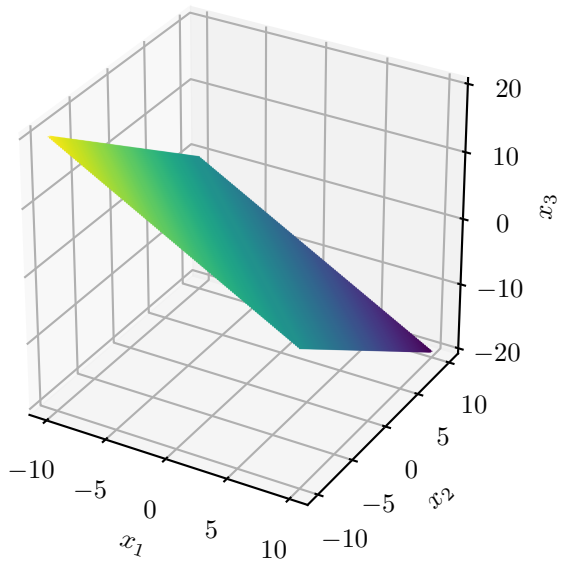

```
import numpy as np
import matplotlib.pyplot as plt

x1 = np.linspace(-10,10)
x2 = 1/2 * x1 - 1

plt.plot(x1,x2)
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
```

b. ($n = 3$ example) Draw the hyperplane for $w = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$, $b = 0$.

This is the equation $x_1 + x_2 + x_3 = 0$, or $x_3 = -x_1 - x_2$.




```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

x1 = np.linspace(-10,10)
x2 = np.linspace(-10,10)
x1grid, x2grid = np.meshgrid(x1,x2)
x3grid = -x1grid - x2grid

fig = plt.figure(constrained_layout=True)
ax = plt.axes(projection='3d')

ax.contour3D(x1grid, x2grid, x3grid, 300)
ax.set_xlabel("$x_1$")
ax.set_ylabel("$x_2$")
ax.set_zlabel("$x_3$")
ax.dist=11
```

c. Given some $x_0 \in \mathbb{R}^n$, find the *squared distance* to the hyperplane defined by $w^T x + b = 0$. In other words, solve the following optimization problem:

$$\min_x \|x_0 - x\|^2$$
$$\text{s.t. } w^T x + b = 0$$

As w is normal to the plane, w and $(x_0 - \tilde{x}_0)$ are parallel or anti-parallel. Thus,

$$|w^T(x_0 - \tilde{x}_0)| = \|w\| \, \|x_0 - \tilde{x}_0\| .$$

Using the constraint above, $w^T \tilde{x}_0 = -b$. Plugging in and rearranging,

$$\|x_0 - \tilde{x}_0\|^2 = \left(\frac{w^T x_0 + b}{\|w\|} \right)^2 .$$

A.10 For possibly non-symmetric $A, B \in \mathbf{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^T A x + y^T B x + c$. Define $\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial z_1} & \frac{\partial f(x,y)}{\partial z_2} & \cdots & \frac{\partial f(x,y)}{\partial z_n} \end{bmatrix}^T$.

a. Explicitly write out the function $f(x, y)$ in terms of the components of A_{ij} and B_{ij} using appropriate summations over the indices.

$$f(x, y) = c + \sum_{i,j} (A_{ij} x_i x_j + B_{ij} y_i x_j)$$

b. What is $\nabla_x f(x, y)$ in terms of the summations over indices *and* vector notation?

$$\text{In index notation:} \quad [\nabla_x f(x, y)]_i = \sum_j A_{ij} x_j + A_{ji} x_j + B_{ji} y_j$$
$$\text{In vector notation:} \quad \nabla_x f(x, y) = (A + A^T) x + B^T y$$

c. What is $\nabla_y f(x, y)$ in terms of the summations over indices *and* vector notation?

$$\text{In index notation:} \quad [\nabla_y f(x, y)]_i = \sum_j B_{ij} x_j$$
$$\text{In vector notation:} \quad \nabla_x f(x, y) = Bx$$

Programming

A.11 For the A, b, c as defined in Problem 8, use NumPy to compute (take a screenshot of your answer):

a. What is A^{-1}?

```
import numpy as np

A = np.array([[0, 2, 4], [2, 4, 2], [3, 3, 1]])
invA = np.linalg.inv(A)

print(invA)
```

```
[[ 0.125 -0.625  0.75 ]
 [-0.25   0.75  -0.5  ]
 [ 0.375 -0.375  0.25 ]]
```

b. What is $A^{-1}b$? What is Ac?

```
import numpy as np

A = np.array([[0, 2, 4], [2, 4, 2], [3, 3, 1]])
invA = np.linalg.inv(A)
b = np.array([-2, -2, -4])
c = np.array([1, 1, 1])

print("A^{-1}b")
print(invA @ b)

print("\nAc")
print(A @ c)
```

```
A^{-1}b
[-2.  1. -1.]

Ac
[6 8 7]
```

A.12 Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y, respectively, are equal, i.e. for all x, $|F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance-$1/k$ variables converges to a (standard) Normal distribution as k goes off to infinity. We will study this phenomenon empirically (you wil use the Python packages Numpy and Matplotlib). Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} B_i$ where each B_i is equal to -1 and 1 with equal probability. From your solution to problem A.5, we know that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$.

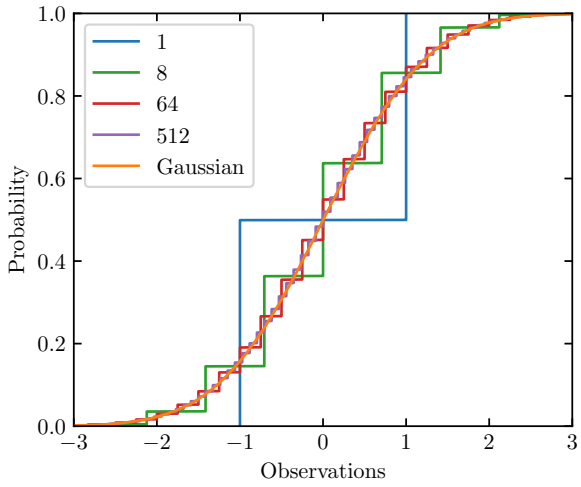
a. For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. If $F(x)$ is the true CDF from which each Z_i is drawn (i.e., Gaussian) and $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x\}$, use the answer to problem A.6 above to choose n large enough such that, for all $x \in \mathbb{R}$, $\sqrt{\mathbb{E}[(\hat{F}_n(x) - F(x))^2]} \leq 0.0025$, and plot $\hat{F}_n(x)$ from -3 to 3.

   Using the solution for A.6, we have

$$n = \frac{F(x)(1 - F(x))}{(0.0025)^2} \leq \frac{1}{4(0.0025)^2} = 4 \times 10^4.$$

   See plot and code below.

b. For each $k \in \{1, 8, 64, 512\}$ generate n independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.




```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(11)

n = int(4e4)

for k in [1,8,64,512]:
        Y = np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1/k), axis=1)
        plt.plot(sorted(Y), np.arange(1,n+1)/float(n), label=k)

Z = np.random.randn(n)
plt.step(sorted(Z), np.arange(1,n+1)/float(n), label='Gaussian')

plt.legend()
plt.xlim(-3,3)
plt.ylim(0,1)
plt.xlabel('Observations')
plt.ylabel('Probability')
```

Homework 0 A

Fall 2020, CSE 546: Machine Learning
John Franklin Crenshaw
October 5, 2020

Probability and Statistics

B.1 Let $X_1, \ldots, X_n$ be n independent and identically distributed random variables drawn uniformly at random from $[0, 1]$. If $Y = \max\{X_1, \ldots, X_n\}$ then find $\mathbb{E}[Y]$.

Let $F_X(x)$ be the CDF and $f_X(x)$ be the PDF of the uniform distribution from which X is drawn. As the distribution is uniform on 0 to 1, $f_X(x) = 1$ in this domain and zero elsewhere. Therefore,

$$F_X(x) = \int_{-\infty}^{x} f_X(x')dx' = \int_0^x dx' = x.$$

Now the probability that the max of $\{X_1, \ldots, X_n\}$ is less than y is

$$F_Y(y) = P(\max\{X_1, \ldots, X_n\} < y) = \prod_{i=1}^{n} P(X_i < y) = \prod_{i=1}^{n} F_X(y) = \prod_{i=1}^{n} y = y^n.$$

Taking the derivative of this CDF, we get the PDF:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = ny^{n-1}.$$

Finally, we can calculate $\mathbb{E}[Y]$:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^1 y \cdot ny^{n-1} = \frac{n}{n+1}$$

B.2 Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathbb{E}[(X - \mu)^2] = \sigma^2$. For any $x \geq 0$, use Markov's inequality to show that $\mathbb{P}(X \geq \mu + \sigma x) \leq 1/x^2$.

Using Markov's Inequality, $\mathbb{P}(X \geq a) = \frac{\mathbb{E}[X]}{a}$, we have

$$\mathbb{P}(X \geq \mu + \sigma x) = \mathbb{P}[(X - \mu)^2 \geq \sigma^2 x^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{\sigma^2 x^2} = \frac{\sigma^2}{\sigma^2 x^2} = \frac{1}{x^2}.$$

Thus, $\mathbb{P}(X \geq \mu + \sigma x) \leq 1/x^2$.

Linear Algebra and Vector Calculus

B.3 The *trace* of a matrix is the sum of the diagonal entries; $\text{Tr}(A) = \sum_i A_{ii}$. If $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$, show that $\text{Tr}(AB) = \text{Tr}(BA)$.

$$\text{Tr}(AB) = \sum_i [AB]_{ii} = \sum_i \left(\sum_j A_{ij} B_{ji} \right) = \sum_j \left(\sum_i B_{ji} A_{ij} \right) = \sum_j [AB]_{jj} = \text{Tr}(BA)$$

B.4 Let $v_1, \ldots, v_n$ be a set of non-zero vectors in $\mathbb{R}^d$. Let $V = [v_1, \ldots, v_n]$ be the vectors concatenated.

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

Let $M = \sum_{i=1}^{n} v_i v_i^T$. As the vectors v_i are in $\mathbb{R}^d$, at most d of them can be linearly independent. Of course, if $n < d$, then at most n vectors can be linearly independent. If you imagine the case where $v_1, \ldots, v_n$ are drawn from an orthonormal basis of $\mathbb{R}^d$, then you can see that $\text{rank}(M) \leq \min(n, d)$. We can also imagine the case where all of the vectors are the same. Then $\text{rank}(M) = 1$. The rank cannot be zero, as the vectors are non-zero. Therefore, $1 \leq \text{rank}(M) \leq \min(n, d)$.

b. What is the minimum and maximum rank of V?

The argument from part a works here as well, just replacing N for V. Thus $1 \leq \text{rank}(V) \leq \min(n, d)$.

c. Let $A \in \mathbb{R}^{D \times d}$ for $D > d$. What is the minimum and maximum rank of $\sum_{i=1}^{n} (Av_i)(Av_i)^T$?

The resultant matrix is $AMA^T \in \mathbb{R}^{D \times D}$, where M is defined above. Despite being a matrix with D columns and rows, it cannot have greater rank than M. This can be seen via the same arguments given in part a, because acting on n vectors in $\mathbb{R}^d$ with the same linear transformation cannot result in more linearly independent vectors than you started with. However, multiplying by A can *reduce* the rank, as the image of A may be lower dimensional than the set of v_i. So $1 \leq \text{rank}(AMA^T) \leq \min(n, d, \text{rank}(A))$.

d. What is the minimum and maximum rank of AV? What if V is rank d?

Again, we can use the same argument as we did in part c. Thus $1 \leq \text{rank}(AV) \leq \min(n, d, \text{rank}(A))$.