Problem 1 (30 Points; Gene Networks). Consider the following RPKM values from 5 RNA-seq experiments for the following genes:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exp1</th>
<th>Exp2</th>
<th>Exp3</th>
<th>Exp4</th>
<th>Exp5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.5</td>
<td>1.3</td>
<td>3.3</td>
<td>4.3</td>
<td>5.7</td>
</tr>
<tr>
<td>B</td>
<td>5.8</td>
<td>6.9</td>
<td>4.5</td>
<td>1.3</td>
<td>1.8</td>
</tr>
<tr>
<td>C</td>
<td>7.8</td>
<td>10.0</td>
<td>15.6</td>
<td>20.9</td>
<td>35.6</td>
</tr>
<tr>
<td>D</td>
<td>8.6</td>
<td>7.0</td>
<td>6.5</td>
<td>7.1</td>
<td>8.7</td>
</tr>
<tr>
<td>E</td>
<td>18.8</td>
<td>14.7</td>
<td>7.5</td>
<td>15.1</td>
<td>18.2</td>
</tr>
</tbody>
</table>

Draw the gene network with the following criteria for edges:

(A) Draw an edge between genes X and Y if the Euclidean distance:

\[ d_E(X,Y) := \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} < 12 \]

. \( n \) is the number of samples (i.e. experiments) and \( X,Y \in \{ A, B, C, D, E \} \).

(B) Draw an edge between genes X and Y if the correlation coefficient \( |r(X,Y)| > 0.8 \). Color the edges with positive correlation red and the edges with negative correlation blue.

(C) Draw an edge between genes X and Y if the \( L_1 \)-norm:

\[ ||X,Y||_{L_1} := \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| < 8 \]

(D) Draw an edge between genes X and Y if the mutual information:

\[ I(X,Y) := \sum_{x \in X} \sum_{y \in Y} p(x,y) \log(\frac{p(x,y)}{p(x)p(y)}) > 0.6 \]

To calculate the mutual information, you bin the RPKM values for each gene into 3 intervals.
Hint: You can use the R package `infotheo` for binning and calculation of the mutual information.

**Problem 2** (40 Points; Probabilistic Distribution, Independence, Information Theory). Consider two random variables $X$ and $Y$ from which we drew the following samples:

$$x = (0.51, 0.99, 0.64, 0.50, 0.27, 0.12, 0.01, 0.79, 0.56, 0.17)$$

$$y = (0.74, 0.06, 0.43, 0.12, 0.61, 0.73, 0.57, 0.91, 0.59, 0.80)$$

First, bin the data by dividing the interval of $[0, 1]$ into 4 equally wide sub-intervals. Provide the following calculations by hand.

(A) Calculate the joint probability distribution $p_{X,Y}(x,y)$ of the binned data and write it in the following table:

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(B) Calculate the marginal distributions $p_X(x)$ and $p_Y(y)$

(C) Calculate the product of the two marginal distributions $p_X(x) \times p_Y(y)^T$ (matrix multiplication!) and compare it with the joint distribution $p_{X,Y}(x,y)$. Are variables $X$ and $Y$ stochastically independent? Justify your answer.

(D) Calculate the conditional distributions $p_{X|Y}(x|y = y_3)$ and $p_{Y|X}(y|x = x_4)$

(E) Calculate the joint entropy $H(X,Y)$ and the marginal entropies $H(X)$ and $H(Y)$

(F) Calculate the conditional entropies $H(X|Y)$ and $H(Y|X)$ using the chain rule.

(G) Calculate the mutual information $I(X,Y)$ using both, the definition and the relation to entropy. Are both results equal? Why?

**Problem 3** (20 Points; Gaussian distribution). Analyze the following two cases of Gaussian distribution.

(A) Consider a two-dimensional random variable $(X,Y)$ from a bivariate Gaussian distribution:

$$(X,Y) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right).$$

Draw a random sample from the bivariate Gaussian distribution with the size $n_1 = 10$ and calculate the correlation coefficient of the two vectors. Repeat the step for sample size $n_2 = 100$ and $n_3 = 1000$. What do you observe? What is the relation of the correlation coefficient and the covariance matrix?
(B) Draw a random sample of size \( n = 100 \) from a univariate Gaussian distribution with mean \( \mu = 0 \) and variance \( \sigma^2 = 3 \). Draw another random sample of the same size from a univariate Gaussian distribution with mean \( \mu = 1 \) and variance \( \sigma^2 = 2 \). Calculate the correlation coefficient of these two vectors. Are the two variables independent?

*Hint*: You can use the R package `MASS` for simulation of the multivariate Gaussian distribution. Use `set.seed(n)` and give the integer \( n \) that you used for your simulation.

**Problem 4** (10 Points; Expected value). Consider the following density function \( f_X(x) \) of random variable \( X \):

![Density function graph]

Calculate the expected value \( EX \).