**Лабораторная работа по основам теории информации**

Работа направлена на закрепление материала по теме «Основные понятия теории информации», а также улучшение навыков владения языком Матлаб.

*Примечание: в Матлабе есть два способа хранения строк:*

1. *Массив символов:* 'abcd'*. Каждый символ – элемент строки:* 'b'=str(2)*. Вертикальной конкатенации подаются только массивы одинаковой длины, то есть полноценный массив строк не реализуется через массивы:*

*~~['abc'; 'de'] % ошибка~~*

1. *Собственно строки:* "abcd"*. Нечто вроде ячейки, которая способна содержать массив символов:* 'abcd'=str{1}, 'b'=str{1}(2). *Поскольку из ячеек можно собирать массивы, можно собирать вертикальные, горизонтальные массивы строк, и даже матрицы строк:*

["abc";"de"]

*Тем не менее,* "abcd" *не эквивалентно* {'abcd'}*: в отличие от ячейки, массив строк может содержать исключительно символьные массивы; также в отличие от ячеек, строки можно сравнивать.*

*Далее по умолчанию будем использовать первый тип, то есть массив символов.*

Работа построена так, что каждый пункт реализуется проще всего использованием функций, написанных ранее.

Как говорилось, истинное распределение вероятностей неизвестно, однако можно вычислить *эмпирические вероятности* встречаемости символов – именно их требуется использовать для подсчёта характеристик величин. Предполагаются стационарность и эргодичность источника.

В файле strings.mat в двух столбцах даны две строки X и Y, содержащие 100 *совместных* реализаций случайных величин X и Y (то есть даны пары (X(i),Y(i)) ). Их удобно использовать при тестировании. Извлечение:

load strings

1. Написать функцию alphabet\_probabilities(), принимающую на вход массив и возвращающую два массива: массив символов алфавита сообщения и соответствующий массив *эмпирических вероятностей* встречаемости символов. Для выделения алфавита удобно воспользоваться встроенной функцией unique(). Не обязательно, но желательно реализовать функцию без использования циклов.
2. Написать функцию entropy(), принимающую на вход массив символов и возвращающую одно число. Функция реализует подсчёт энтропии (первого порядка, т.е. безусловной) символа во входной строке. Не использовать циклы. *Подсказка: можно использовать в реализации массив частот, получаемый предыдущей функцией*.
3. Написать функцию cond\_val\_entropy(), которая принимает на вход массивы X и Y, а также значение символа y и возвращает одно число. Функция, если среди значений Y встречается y, считает условную энтропию при данном значении: H(X|Y=y). Без использования циклов *(применить логическую индексацию)*.
4. Написать функцию cond\_entropy(), которая принимает на вход массивы X и Y и возвращает одно число, равное условной энтропии H(X|Y). Можно использовать цикл.

Написать функцию joint\_entropy(), которая принимает на вход массивы X и Y и возвращает одно число, равное совместной энтропии H(X,Y). Можно использовать цикл. *Подсказка:* string(['ab';'cd']) = ["ab", "cd"]*. Верно применяя данное преобразование и транспонирование, удобно получить массив из строк, содержащих по два символа (по одному из X и Y). Эти строки можно сравнивать между собой, следовательно, ранее написанные функции окажутся применимыми к ним.*

1. Для X и Y из файла посчитать H(X), H(Y), R(X), R(Y), H(XY), H(Y|X), H(X|Y), I(X;Y), I(Y;X), проверить равенства и неравенства:
2. В лекции мы определяли энтропию источника через предел двух величин: условной L-энтропии:

И средней L-энтропии:

Оценить энтропию источника через условную L-энтропию и среднюю L-энтропию для L=…. Сравнить с энтропией символа и энтропией нулевого порядка.  
*Подсказка*: для индексации понадобится такая матрица размера :

Она формируется таким образом:

(1:L)+(0:n-L).'

Также пригодится подсказка из пункта 4