Information Theory and Applications

Assignment 2

Prof. Giorgio Taricco

Politecnico di Torino – DET

Assignment rules

- No cooperation outside of the group is allowed
- The report must describe all the details of the code, the numerical results in a readable form, and the derivation of the analytic parts when required
- The report must be well written and organized, it must provide all the implementation details, and its quality is an important part of the evaluation
- The report must contain pointers to the code developed (for example, this result has been obtained by using the Matlab function xxx available in the assignment delivery package) and all the assumptions made
- Matlab is the preferred implementation language
- Notebook Python is accepted with detailed code description and pointers in the report plus comments in the notebook

1 – Basic notions

- For each exercise provide the analytic solution (if it is possible to find an analytic solution, and include the parameter ranges)
- Otherwise, state that the analytic solution is impossible and provide the numerical solution (the correct choice is part of the evaluation)
- Calculate the entropy of the following distributions:
    1. $p(n) \propto \exp(-\lambda n)\,, n = 0,1,2,\dots$
    2. $p(n) \propto \exp(-n^2)\,, n = 0,1,2,\dots$
    3. $p(n) \propto n^{-4}, n = 1,2,3,\dots$
    4. $p(n) \propto \alpha^n, n = 1,\dots,N$
    5. $p(n) \propto (1+n^2)^{-k}, n = 0,1,2,\dots$ for $k = 1,2$
- Calculate the differential entropy of the following pdfs:
    1. $f_X(x) = \frac{u(x-a)-u(x-b)}{b-a}$
    2. $f_X(x) = \lambda e^{-\lambda x} u(x)$
    3. $f_X(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x} u(x)$

2 – Markov source

- A Markov source with alphabet $\{1,2,3,4,5\}$ is characterized by the following conditional probabilities:
$$P(X_n = x_n | X_{n-3:n-1} = x_{n-3:n-1}) \propto (\max\{x_{n-3:n-1}\})^{x_n/5}$$
- Calculate the asymptotic state distribution of the source
- Calculate the entropy rate

CAVEAT: when a probability distribution is specified by proportionality (∝), normalization must be applied!

3 – Huffman code

- Given the stationary source with probabilities

$$p(n) \propto \frac{1}{1+n^3}, \qquad n = 0,1,\dots,N-1$$

- Consider the following three cases: $N = 10,100,1000$ and, for each case,
  - Calculate the entropy rate
  - Calculate the average number of bits per symbol required by a Huffman code

4 – Text file encoding

- Build a dictionary with the words from the file “warandpeace.txt” and the number of occurrences of each word.
- Calculate the entropy rate by using the word frequencies as probabilities.
- Apply Huffman encoding to encode the N words with the higher frequencies.
- Encode the remaining words by maximum-length coding.
- Complete the details of the encoder.
- Evaluate the memory occupied by the dictionary specifying the storage format.
- Evaluate the memory occupied by the encoded text as a function of N