Homework 10-11: On source coding

CR07: Selected Topics Information Theory (Fall 2019). ENS de Lyon
jean-marie.gorce@inria.fr – Deadline: December, 23, 2019 at 11:59 p.m.

1 Lossless compression

1.1 Preliminaries

Consider a source $X = S^n$ where S_i are identically and independently distributed on an alphabet $\mathcal{A}$, where $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$, where a_k have the following probabilities $P(a_k) = \beta \cdot exp(-\alpha k)$, where $\alpha \in \mathbb{R}^+$ and and β is a normalization constant appropriatly chosen.

We note $f^*(X)$ the optimal variable-length compressor. This means that the length $l(f^*(X))$ is stochastically dominated by $l(f(X))$ for any other compressor $f(X)$.

1.2 Questions

1. Compute the entropy and the varentropy of this source.
2. Compute and plot the exact cumulated distribution function of the length of the optimal compressor, for different values of n.
3. Compare to the asymptotic Gaussian approximation and to the upper and lower bounds derived in the course.
4. Compare the average length to the bounds on average length.
5. Analyse the results.

2 Almost lossless compression

2.1 Preliminaries

Consider a gray level image where each pixel is coded on 8 bits, with a value between 0 and 255.

We consider 2 situations, independently :

- Grey levels are uniformly distributed.
- Grey levels are exponentially distributed as in the former problem.

2.2 Questions

1. Compute the compression factor achievable with a variable length compressor as a function of the image size for the two former situations, independently.
2. Compute and compare the achievable compression factor with uniquely decodable codes.
3. Discuss the performance of the Huffman code applied to each pixel independently.
4. Prove that Huffman code is an optimal uniquely decodable code when it is applied to the source X as a whole.
5. Evaluate the performance when the Huffman code is applied for groups of pixels, and analyse the complexity-performance tradeoff.

3 Fixed length code, almost lossless

3.1 Preliminaries

We consider now the same problem, for which we want to perform a fixed length coding.

3.2 Questions

1. In the case of the exponential distribution of the grey levels, compute and compare the optimal compressor-decompressor pairs under non detectable error and under detectable error.

2. Compare the performance and the advantages/drawbacks of fixed lendth compressors versus variable-length compressors.

4 Slepian-Wolf

4.1 Preliminaries

The Slepina-Wolf compression is defined as the compression with side information at the receiver only. The minimal error of this code is bounded as follows:

$$\epsilon^*(X|Y,k) \leq \epsilon^*_{SW}(X|Y,k) \leq \mathbb{P}\left[i_{X|Y}(X|Y) \geq k-\tau\right] + 2^{-\tau} \tag{1}$$

5 Lossy compression

5.1 Preliminaries

We now consider the lossy compression problem. consider the former image compression problem. We still assume that the pixels are i.i.d.

We want to compress this image to only 4 bits per pixel.

5.2 Questions

1. What is the optimal scalar quantizer ?
2. What is the maximal error gain we can expect with a vector quantizer ? Give a theoretical proof.
3. Give the fundamental rate-distorsion limit for this problem.
4. Explain how correlation between pixels may help to further compress with the same error rate.