Source Coding

Introduction Diagram of a general communication system. *Discrete sources* output of the source is in discrete time and discrete valued. *Source Coding* representation of information sources in bits. *Source Code Function* $C : U \mapsto \{0,1\}^* = \{\emptyset, 0, 1, 00, ...\}$.
Non-Singular Codes A code C is *singular* if $\exists u \neq v / \ C(u) = C(v)$. A code C is *non-singular* if it is not singular. With a code C define for a positive integer $n : \ C^n : U^n \mapsto \{0,1\}^*$ as $C^n(u_1, u_2, ..., u_n) = C(u_1)C(u_2)...C(u_n)$ $C^* : U^* \mapsto \{0,1\}^*$ as $C^*(u_1 u_2 ... u_n) = C(u_1)C(u_2)...C(u_n)$
Uniquely Decodable Codes A code C is said to be *uniquals decodable* if C^* is non-singular.We want our codes to be uniquals decodable.
Prefix-Free Codes A sequence $u_1, ...u_n$ is a *prefix* of $v_1, ..., v_n$ if $n \geqslant m / \ u_1 = v_1, ..., u_m = v_m$. A code C is said to be *prefix-free* if $\forall u \neq v \ C(u)$ is not a prefix of $C(v)$.
Theorem A prefix-free code is uniquely decodable. (In a binary-tree representation of a PF code all codewords are found on the leaves).
Kraft's Inequality for PF Codes *Theorem* If C is PF then $Kraftsum(C) \triangleq \sum_{u \in U} 2^{-length(C(u))} \leq 1$. *Proposition* $Kraftsum(C^n) = [Kraftsum(C)]^n$.
Kraft's Inequality for extensions of codes *Proposition* Suppose $C : U \mapsto \{0,1\}^*$ is a non-singular code then $Kraftsum(C) = \sum_{u \in U} 2^{-length(C(u))} \leq 1 + max\left[length(C(u))\right]$
Kraft's Inequality for uniquely decodable codes *Theorem* If C is a uniquely decodable code then $Kraftsum(C) \leq 1$. *Corollary* If C is a uniquely decodable code then there exists a PF code C' such that $length(C(u)) = length(C'(u))$.
Reverse Kraft's inequality *Theorem* Given an alphabet U and a function $l : u \mapsto \{0,1,2,3,...\} / \ \sum_{u \in U} 2^{-length(C(u))} \leq 1$ then there exist a PF code $C : U \mapsto \{0,1\}^* / \ \forall u \in U$ $length(C(u)) = l(u)$
Sources A source producer a sequence $u_1, u_2, u_3, ...$ each $u_i \in U$ being random variables. A *memory-less* source is one where $u_1, u_2, ...$ are independent. A *stationary* source is one where each $(u_i, ..., u_{i+n-1})$ has the same statistics as $(u_1, ..., u_n)$ for each i and each n. A memory-less and stationary source is equivalent to $u_1, u_2, ...$ are *independent, identically distributed (iid)*.
Expected codeword length $E\left[length(C(u))\right]$ average number of bits/letter the code uses to represent the source. We want to minimize it and C to be uniquely decodable.

Entropy

Lemma $ln(z) \leq z^{-1}$ with eq if $z = 1$. *Property* $0 \leq H(U) \leq log|U|$ **Entropy as a lower-bound to the expected codeword length** *Theorem* For any uniquely decodable code C for a source U, we have $E\left[length(C(u))\right] \geq \sum_u p(u) log_2 \frac{1}{p(u)} \triangleq H(u)$
Existence of PF codes with average length at most entropy + 1 *Theorem* Given source U there exists a PF code C s.t. $E\left[length(C(u))\right] \geq H(u) + 1$
Entropy of multiple random variables *Property* Suppose U and V are ind. RV. Then $H(UV) = H(U) + H(V)$. *Observe* Suppose we have $U_1 U_2 ...$ iid. If we use a code C to represent n letters at time., we will have $H(U_1 ... U_n) \leq E\left[length(C(U_1 ... U_n))\right] \leq H(U_1 ... U_n) + 1$. *Also* $\frac{1}{n} H(U_1 ... U_n) = H(U_1)$ (iid of U).
Properties of optimal codes *1* If $p(u) < p(v)$ then $l(u) \geq l(v)$. *2* In an optimal PF code there are more than 2 longest codewords. If not the longest codeword can be shortered without violating the PF condition. *3* Among optimal codes, there is one for the two least probable symbols are siblings.
Huffman procedure Procedure to design the optimal code. *1* Given prob $p_1, p_2, ..., p_{k-1}, p_k$. Start with the two smallest prob. *2* Group them together as the binary descendant of a node. *3* Repeat until one node is left.
Equivalence of PF codes and strategy for guessing via binary questions TODO
Interpretation of entropy as expected number of questions for guessing the random variable TODO

Mutual Information

Conditional Entropy and Mutual Information *Conditional Entropy* $H(U|V = v) = \sum_u p(u|v) log \frac{1}{p(u|v)}$ $H(U|V) = \sum_v p(v) H(U|V = v)$. *Conjecture* $H(U|V) \leq H(U)$. *Mutual Information* $I(U;V) = H(U) + H(V) - H(UV)$ is the saving in the number of questions to given U by the knowledge of V. *Lemma* Suppose W is an alphabet and p and q are two prob distribution in W. Then, $\sum_w p(w) log \frac{p(w)}{q(w)} \geq 0$ with eq. iff $p = q$. *Theorem* $I(U;V) \geq 0$ with eq iff U and V are independent. *Conditional Mutual Information* $I(U;V|W) = H(U|W) + H(V|W) - H(UV|W) = H(U|W) - H(U|VW) = H(V|W) - H(V|UW)$ *Theorem* $I(U;V|W) \geq 0$ with eq. iff U,V are independent conditional in $W \equiv U - V - W$.
Chain Rules for entropy and mutual information *Theorem* $H(UV) = H(U) + H(V|U) = H(V) + H(U|V)$ *Theorem* $H(U_1 ... U_n) = H(U_1) + H(U_2|U_1) + ... + H(U_n|U_1 ... U_{n-1})$ *Theorem* $I(U_1 ... U_n, V) = I(U_1;V) + ... + I(U_n;V|U_1 ... U_{n-1})$
Review of Markov Chain Suppose X,Y,Z are RVs. We can write $p(xyz) = p(x)p(y|x)p(z|xy)$. Because $p(y|x) = \frac{p(xy)}{p(x)}$. If $X - Y - Z$ then $p(xyz) = p(x)p(y|x)p(z|y)$. Suppose $U_1 - U_2 - ... - U_n$ then $H(U_1 ... U_n) = H(U_1) + H(U_2|U_1) + ... + H(U_n|U_{n-1})$
Data Processing Inequality *Theorem* Suppose $U - V - W$ then $I(U;W) \leq I(U;V)$ *Corollary* If $U - V - W$ then $I(U;W) \leq I(V;W)$. *Corollary* If $U - V - W - X$ then $I(U;X) \leq I(V;W)$. $I(UV;W) = I(U;W) + I(V;W|U) = I(V;W) + I(U;W|V)$.
Entropy Rate Given a stochastic process $U_1, U_2 ...$ we define its *entropy rate* $H(U) = \lim_{n \to \infty} \frac{1}{n} H(u_1 ... u_n)$ if the limit exists.
Entropy Rate of Stationary Processes *Theorem* If $u_1, u_2, ...$ is stationary process, then the entropy rate exists and $\lim_{n \to \infty} \frac{1}{n} H(u_1 ... u_n) = \lim_{n \to \infty} H(u_n|u_1 ... u_{n-1})$
Coding Theorem for Stationary Sources *Theorem* If $U_1, U_2, ...$ is a stationary process with entropy rate H, then $\forall \varepsilon > 0$ there exists a source code $C_n : U^n \to \{0,1\}^*$ s.t. the average code length is less than $H + \varepsilon$
Fixed-to-Fixed Length Source Codes Codes of type $U \to \{0,1\}^*$ or $U^n \to \{0,1\}^*$

Typicality

Typicality TODO
Properties of Typical Sets TODO
Asymptotic Equipartition Property TODO
Source Coding by AEP TODO
Variable-to-Fixed Length Source Codes TODO
Valid and Prefix-Free Dictionaries TODO
Relationship between word- and letter-entropies for valid, prefix-free dictionaries TODO

Tunstall procedure

Tunstall procedure TODO
Analysis of Tunstall procedure TODO
Universal Source Coding TODO
LempelZiv method TODO
Analysis of LempelZiv TODO
Information-Lossless FSM Compressors TODO
Lower bound on the output length of an IL FSM Compressor TODO
LZ Compressibility of sequences TODO
Optimality of LempelZiv TODO

Channels

Communication Channels TODO
Discrete Memoryless Channels TODO
Examples of Discrete Memoryless Channels (BSC and BEC) TODO
Transmission with or without feedback TODO
Channel Capacity TODO
Fano's Inequality TODO
Converse to the Channel Coding Theorem TODO
Proof of the Channel Coding Theorem TODO
Capacity of BSC and BEC TODO
Jensen's Inequality TODO
Concavity of Mutual Information in Input Distribution TODO

KKT Conditions TODO

KKT Conditions (cont'd) TODO

Application of KKT: Capacity of Z Channel TODO

Continuous Alphabet: Differential Entropy TODO

Properties of differential entropy TODO

Entropy-typical seqeuences TODO

Quantization TODO

Entropy of Gaussian distribution TODO

Capacity under cost constraint TODO

Capacity of AWGN TODO

Converse to the channel coding theorem with cost constraint TODO

Parallel Gaussian channels (water-filling) TODO

Proof of Channel Coding Theorem for general channels via Threshold Decoding TODO

Channel Codes TODO

Minimum Distance TODO

Singleton Bound TODO

Sphere-packing Bound TODO

GilbertVarshamov Bound TODO

Linear Codes TODO

Generator Matrix TODO

Parity-check Matrix TODO

Hamming Codes TODO

ReedSolomon Codes TODO

Polar Codes TODO

Credits

Most content taken from the lecture notes of Emre Telatar's Information Theory and Coding class at EPFL, 2015.

Rendered December 30, 2015. Written by Lucile Madoulaud.