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 $C^*(u_1u_2...u_n) = C(u_1)C(u_2)...C(u_n)$
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 $Kraftsum(C) = \sum_{u \in U} 2^{-length(C(u))} \leq 1 + max\,[length(C(u))]$
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 $E\,[length(C(u))]$ 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 $E\,[length(C(u))] \geq \sum_u p(u) log_2 \frac{1}{p(u)} \triangleq H(u)$
Existence of PF codes with average length at most entropy + 1 *Theorem* Given source U there exists a PF code C s.t. $E\,[length(C(u))] \geq H(u) + 1$
Entropy of multiple random variables *Property* Suppose U and V are ind. RV. Then $H(UV) = H(U) + H(V)$. *Observe* Suppose we have $U_1U_2...$ iid. If we use a code C to represent n letters at time., we will have $H(U_1...U_n) \leq E\,[length(C(U_1...U_n))] \leq H(U_1...U_n) + 1$. *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 $U^n \to \{0,1\}^*$ are called *fixed-to-variable* length codes, and all our designs have error free recovery of the source from its representation. We want *Fixed-to-fixed* codes $C : U^n \to \{0,1\}^k$ (2^k representations), to obtain efficient codes we will give up error free recovery replace this by recovery with very small prob. of error. The code assign binary representations only to a subset $S \subset U^n$ which ensure $Pr((u_1...u_n) \in S) \approx 1$ and $|S| \leq 2^k$.

Typicality

Typicality Given an alphabet U and distribution p in U. We have a source that produces iid letters u_i with distribution p. We want a *set* $T_{n,\varepsilon} \subset U^n$.
Properties of Typical Sets *1.* $Pr((u_1...u_n) \in T_{n,\varepsilon,p}) \approx 1$ $(\geq 1 - \frac{|U|}{n\varepsilon^2 p(u)})$. *2.* $|T_{n,\varepsilon,p}| \leq 2^{n(1+\varepsilon)H(u)}$
Asymptotic Equipartition Property *Def* AEP is a general property of the output samples of a stochastic source. It is fundamental to the concept of typical set used in theories of compression. *Theorem* If $U_1, ...U_n$ iid $\sim q$ and $(u_1...u_n) \in T_{n,\varepsilon,p}$ then $Pr((U_1...U_n) = (u_1...u_n)) = \prod_{i=1}^n q(u_i)$. *Corollary* $Pr(((U_1...U_n) = (u_1..u_n)) \in T_{n,\varepsilon,p}) = 2^{-n(1 \pm \varepsilon)D(p||q)} 2^{\pm n\varepsilon H(p)} (1 \pm \varepsilon) \doteq 2^{-nD(p||q)}$. *Consequently* If $(X_1,Y_1)...(X_n,Y_n)$ iid $p_X p_Y$ then $Pr(((X_1,Y_1)...(X_n,Y_n)) \in T_{n,\varepsilon,p_{XY}}) \doteq 2^{-nI(X;Y)}$
Source Coding by AEP *1.* Assign to every member of $T_{n,\varepsilon,p}$ a distinct element of $\{0,1\}^*$. Call this $C : T_{n,\varepsilon,p} \to \{0,1\}^*$. *2.* The source code is the following :Observe if $(U_1...U_n) \in T_{n,\varepsilon,p}$ represent it by 0 else by 1.
Variable-to-Fixed Length Source Codes $\equiv$ Dual of Fixed-to-Variable length source coding $\equiv$ Dictionary to sed source coding *Idea* Given an alphabet U, find a dictionary $D \subset U^*$, assign $\lceil log|D| \rceil$ bit binary representation to words in D, and then given $U_1U_2..$, parse it into $w_1, w_2...$ of dictionary words and represent each words by its binary description.
Valid and Prefix-Free Dictionaries The words of valid and PF dictionaries form the leaves of a complete dictionary tree.
Relationship between word- and letter-entropies for valid, prefix-free dictionaries nbr of bits/letter $= \frac{m\lceil log|D| \rceil}{length(W_1) + ... + length(W_m)} \to \frac{\lceil log|D| \rceil}{E[length(W_1)]}$ *Lemma* Suppose Z is non negative integer valued RV. Then $E[Z] = \sum_{n=0}^{\infty} Pr[Z > n]$. *Observation* nbr of words = nbr of leaves in tree representation $= 1 + (|U| - 1)$(nbr interior nodes).

Tunstall procedure

Tunstall procedure *1.* $D = U$ with $|U|$ words and one interior node (root). *2.* While $|D| < M$ do : convert the most probable word/leaf into interior node and grow $|U|$ leaves from it.
Analysis of Tunstall procedure *Lemma* Suppose D valid and PF dictionary and $U_1U_2...$ are iid. Then $H(W_1) = E[length(W_1)]H(U_1)$. *Now* nbr of bits/letter $= \frac{\lceil log|D| \rceil}{E[length(W_1)]} = \frac{\lceil log|D| \rceil}{H(W_1)} H(U_1)$.
Universal Source Coding *Lemma* If W is a RV taking values in D and for each $w \in D$ $qp_0 \leq Pr(W = w) \leq q$. Then $\lceil log|D| \rceil - log\frac{1}{p_0} \leq H(W) \leq \lceil log|D| \rceil$. *Corollary* Given U and p_U, $\varepsilon > 0$ there exists a dictionary based code which $nbrofbits/letter \leq H(U)(1 + \varepsilon)$.
LempelZiv method - Data compression *0.* Set $D = U$ *1.* Associate to each $w \in D$ a $\lceil log|D| \rceil$bit binary representation based on dictionry order. *2.* Parse the next word w from the source sequence using D, emit the representation of w. *3.* Set $D \leftarrow (D \setminus \{w\}) \cup \{wu : u \in U\}$. *4.* Go to 1.
Analysis of LempelZiv *Technique* Compare LZ to an adversary : adversary knows before have $u_1u_2...$. designs a FSM to compress this sequence. And show LZ does as well as the adversary.
Information-Lossless FSM Compressors A *Finite-State-Machine* is 1. Set S of states $|S| < \infty$. 2. Initial special state $s_0 \in S$. 3. Next state function $g : SxU \to S$. 4. $f : SxU \to \{0,1\}^*$. A *legitimate* machine has to verify : $\forall s \in S \forall u_1...u_m \neq v_1...v_n$ if $g(s, u_1...u_m) = g(s, v_1...v_m) \Rightarrow f(s, u_1...u_m) \neq f(s, v_1...v_n)$. *IL* A legitimate FSM is information-lossless.
Lower bound on the output length of an IL FSM Compressor *Fact* Suppose m numbers $a_1...a_m \geq 0$ with $\sum_{i=1}^m a_i = k$. Then, $\sum_{i=1}^m a_i log\frac{a_i}{8} \geq klog\frac{k}{8m}$ (with eq. for $a_i = \frac{k}{m}$).

LZ Compressibility of sequences *Corollary* If $u_1u_2...$ is a stationary process with entropy rate H then $E[p_{LZ(U_1U_2...)}] \le H$.
Optimality of LempelZiv TODO

Channels

Communication Channels *Def* $P_{e,i} = Pr(U_i \neq V_i)$. $\bar{P}_e = \frac{1}{L}\sum_{i=1}^{L} Pr(U_i \neq V_i)$
Discrete Memoryless Channels A channel is said to be *memoryless* if $Pr(Y_i = y|X_i = x_i, past) = Pr(Y_i = y|X_i = x_i)$
Encoder function $f : 1,...,M \to \mathcal{X}^n$.
Decoder function $\Phi : \mathcal{Y}^n \to \{1,...,M\}$.
Recall $h_2(\alpha) = \alpha log(\frac{1}{\alpha}) + (1-\alpha)log(\frac{1}{1-\alpha})$.
Theorem "Bad news" No matter how the encoder and decoder are designed, we have $h_2(\bar{P}_e) + P_e log(|U|-1) \ge \tau_s(\frac{H}{\tau_s} - \frac{C}{\tau_c})$. Or if $R > C$ we show that any design with $rate \ge R$, we can't make bot error prob closer to 0 then some $\delta(C,R) > 0$.
Lemma $h_2(\frac{1}{L}\sum_{i=1}^{L} p_i) \ge \frac{1}{L}\sum_{i=1}^{L} h_2(p_i)$
Rate $rate(f) = R = \frac{k}{n} \equiv R = \frac{logM}{n}$
Probability of error $P_{e,i} = Pr(\Phi(Y_1...Y_n) \neq i|(X_1...X_n) = f(i))$.
Average prob. of error $P_{e,avg} = \frac{1}{M}\sum_{i=1}^{M} P_{e,i}$.
Maximal prob. of error $P_{e,max} = max_{1\le i\le M} P_{e,i}$.
Examples of Discrete Memoryless Channels (BSC and BEC) TODO
Transmission with or without feedback TODO
Channel Capacity $C = max_{P_X} I(X;Y)$ with $I(X;Y) = H(X) - H(X|Y)$.
Theorem Given a channel $P(y|x)$ with $C = max I(X;Y)$ then any $R < C$ is achievable.
Fano's Inequality Suppose U and V RVs taking values in set $\mathcal{U}$. Let $P_{e,i} = Pr(U_i|V_i)$, then $H(U_i|V_i) \le h_2(P_{e,i}) + P_{e,i}log(|\mathcal{U}|-1)$.
Theorem "Good news" If $\frac{H}{\tau_s} < \frac{C}{\tau_c}$ can achieve $\bar{P}_e \to 0$. Or if $R < C, \varepsilon > 0$ then there exists an Enc/Dec s.t. $rate \ge R, P_{e,max} < \varepsilon$
Converse to the Channel Coding Theorem TODO
Proof of the Channel Coding Theorem TODO
Capacity of BSC and BEC *Binary Symmetric Channel* $\mathcal{X} = \mathcal{Y} = \{0,1\}$.
$p(y|x) = \{ \begin{matrix} 1-\delta & ,x=y \\ \delta & ,else \end{matrix}$ $I(X;Y) = H(Y) - H(Y|X) = H(Y) - h_2(\delta) \le 1 - h_2(\delta)$. If we choose $p_X = (\frac{1}{2},\frac{1}{2})$ then $p_Y = (\frac{1}{2},\frac{1}{2})$. So $C = 1 - h_2(\delta)$ bits.
Binary Erasure Channel $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,e\}$. $p(y|x) = \{ \begin{matrix} 1-\varepsilon & ,x=y \\ \varepsilon & ,y=e \end{matrix}$
$I(X;Y) = H(X) - H(X|Y) = (1-\varepsilon)H(X) \le 1-\varepsilon$. If we choose $p_X = (\frac{1}{2},\frac{1}{2})$. So $C = 1-\varepsilon$ bits.

Concavity

Recall convexity concavity f is *convex* if $\forall x_1 \in D, \forall x_2 \in D, \forall\theta \in [0,1]$ $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2)$.
f is *concave* if $\forall x_1 \in D, \forall x_2 \in D, \forall\theta \in [0,1]$ $f(\theta x_1 + (1-\theta)x_2) \ge \theta f(x_1) + (1-\theta)f(x_2) \Leftrightarrow -f$ is convex.
f is *linear* if $\forall x_1 \in D, \forall x_2 \in D, \forall\theta \in [0,1]$ $f(\theta x_1 + (1-\theta)x_2) = \theta f(x_1) + (1-\theta)f(x_2)$.
Properties 1. If f_1, f_2 convex and C_1, C_2 non negative then $f(x) = C_1f_1(x) + C_2f_2(x)$ convex. 2. Any local minimum is a global minimum. 3. Tangent line lies below function. 4. g convex and h linear $\Rightarrow g(h(x))$ convex.
Jensen's Inequality *Theorem* If f is convex and X anx RV on D then $f(E[X]) \le E[f(X)]$. (idem concavity $\ge$).
Concavity of Mutual Information in Input Distribution *Theorem* Fix $p(y|x)$, $I(p_X) = I(X;Y)$ is concave.
KKT Conditions *Def* $(P_1,...,P_k)$ satisfies KKT if $\exists\lambda$ s.t. $\forall i, \frac{\partial f}{\partial p_i} = \lambda$ for $p_i > 0$ and $\frac{\partial f}{\partial p_i} \le \lambda$ for $p_i = 0$.
Theorem For concave differentiable $f(P_i)$, $(P_1,...,P_k) \in argmax_{\sum P_i=1, P_i\ge0} f(P_1,...,P_k)$ iif it satisfies the KKT conditions.
KKT Conditions (cont'd) *Theorem* $p(x) \in argmax I(X;Y)$ iif $\exists\lambda$ s.t. $\forall x \sum_y p(y|x)log(\frac{p(y|x)}{p(y)}) \le \lambda$. With eq. for $p(x) > 0$, if so then $C = \lambda$.
Application of KKT: Capacity of Z Channel *Capacity*
$C = log(1 + 2^{-\frac{h_2(\delta)}{1-\delta}}) = \lambda$
$p_x^*(0) = 1 - \frac{1}{(1-\delta)(1+2^{-\frac{h_2(\delta)}{1-\delta}})}$
Continuous Alphabet: Differential Entropy
$h(X)$ $= E[-log f_X(X)] = -\int f_X(x)log(f_X(x))dx$.
$h(Y|X)$ $= -E[log f_{X|Y}(X|Y)]$.
$h(X,Y)$ $= E[-log f_{XY}(X,Y)]$.
$D(f||g)$ $= E_f[log\frac{f_X(X)}{g_X(X)}] = \int f_X(x)log\frac{f_X(x)}{g_X(x)}dx$.
$I(X;Y)$ $=$
$D(f_{XY}||f_Xf_Y) = \int\int f_{XY}(x,y)log\frac{f_{Y|X}(y|x)}{f_Y(y)}dxdy$
Properties of differential entropy
1. $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$.
2. $D(f||g) \ge 0$ with eq for $f = g$.
3. $I(X;Y) \ge 0$ with eq for X and Y indepdt.
4. Conditioning reduces h: $h(X|Y) \le h(X)$ with eq for X and Y indepdt.
5. Chain Rule : $h(X_1,...,X_n) = h(X_1) + h(X_2|X_1) + ... + h(X_n|X_1...X_{n-1})$.
6. $h(X_1,...,X_n) \le \sum_{i=1}^{n} h(X_i)$.
Entropy-typical sequences *Discrete* $|T^n(\varepsilon)| = 2^{n(H(X)+\delta(\varepsilon))}$. *Continuous* $Vol(T^n(\varepsilon)) = 2^{n(h(X)+\delta(\varepsilon))}$. $P[X \in T^n(\varepsilon)] \xrightarrow[n\to\infty]{} 1$.
Theorem $\lim_{\Delta\to0}(H(X_\Delta) + log\,\Delta) = h(X)$.
Quantization *Theorem* $I(X_\Delta);Y_\Delta) \xrightarrow[\Delta\to0]{} I(X;Y)$

Entropy of Gaussian distribution *Gaussian X* We have $\mathcal{X} \sim N(0,\sigma^2)$ and $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$. Then $h(X) = E[-log_e f_X(X)] = \frac{1}{2}log_e(2\pi e\sigma^2)Nats = \frac{1}{2}log_2(2\pi e\sigma^2)Bits$. *Theorem "Maximal entropy"* If $Var[X] = \sigma^2$, then $h(X) \le \frac{1}{2}log(2\pi e\sigma^2)$ with eq for $X \sim N(\mu,\sigma^2)$ for some μ.
Capacity under cost constraint $C(P) = max_{f_X:E_f[b(X)]\le P} I(X;Y)$
Capacity of AWGN : Additive White Gaussian Noise $Y = X + Z$, $Z \sim N(0,\sigma^2)$. $C(P) = \frac{1}{2}log(1+\frac{P}{\sigma^2})$
Converse to the channel coding theorem with cost constraint *Theorem* If $\frac{H}{\tau_s} > \frac{C(P)}{\tau_c}$, (any code with $\sum_u[p(u)\frac{1}{n}\sum_{i=1}^{n} b(x_i^u \le P))$ has a bit error rate $\bar{P}_e \not\to 0$ which can't be arbitrarily small.
Lemma $\sum_{i=1}^{n} I(X_i;Y_i) \le nC(P)$.
Parallel Gaussian channels (water-filling) With KKT $\exists\lambda...$ so $P_i = max\{0, \lambda - \sigma_i^2\}$. Choose λ so that $\sum_{i=1}^{k} P_i = P$. Indeed, λ : water-level and shaded areas : P_i.
Proof of Channel Coding Theorem for general channels via Threshold Decoding *Theorem* $\forall R < C(P), \forall\varepsilon > 0, \exists n\exists code$ with $\forall m, \frac{1}{n}\sum_{i=1}^{n} b(X_i^{(n)}) \le P$ and $P_{max} \le \varepsilon$.
$P_{max} = max_m P[error|m]$.
$P_{avg} = \frac{1}{M}\sum_{m1}^{M} P[error|m]$.
Corollary If $\frac{H}{\tau_s} < \frac{C(P)}{\tau_c}$ can achieve $\bar{P}_e \to 0$.

Distances

Channel Codes Code for *BSC*: $M = 2^{nR}$. Given the code we see (n,nR) as visible parameters.
Minimum Distance *Hamming distance* $d_H(\bar{x},\bar{y}) = \#\{i : y_i \neq x_i\}$ so maximum-likehood $\equiv$ minimum distance. *minimum distance of C to these parameters* $d = d_H(C) = min_{x,x'\in C, x\neq x'} d_H(x,x')$.
Hamming ball with center $\bar{x}$ and radius r $:B(\bar{x},r) = \{\bar{y} \in \{0,1\}^n : d_H(\bar{y},\bar{x}) \le r\}$. Also note $|B(\bar{x},r)| = 1 + \binom{n}{1} + ... + \binom{n}{r}$
Theorem Given $\bar{x},\bar{y},\bar{z} \in \{0,1\}^n$. $d_H(\bar{x},\bar{y}) \ge 0$ with eq if $\bar{x} = \bar{y}$. $d_H(\bar{x},\bar{y}) = d_H(\bar{y},\bar{x})$. $d_H(\bar{x},\bar{z}) \le d_H(\bar{x},\bar{y}) + d_H(\bar{y},\bar{z})$.
Corollary Suppose $\bar{x},\bar{x'} \in \{0,1\}^n$ with $d = d_H(\bar{x},\bar{x'})$. Set $r = \lfloor\frac{d-1}{2}\rfloor$. Consider $B(\bar{x},r)$ and $B(\bar{x'},r)$ then they are disjoint.
Singleton Bound (Bad news) If $\#ofcodewords = M > 2^k$ then $d_{min} \le n-k$.
Sphere-packing Bound (Bad news) Suppose $C \subset \{0,1\}^n$ is a binary code with M codewords and $d = d_{min}$. Then, $M\left[\sum_{i=0}^{r=\lfloor\frac{d-1}{2}\rfloor}\binom{n}{i}\right] \le 2^n$.
GilbertVarshamov Bound (Good news) For every (n,M,d) satisfy $d_{min} = d$ and $M\left[\sum_{i=0}^{d-1}\binom{n}{i}\right] \ge 2^n$ there exists a code $C \subset \{0,1\}^n$ with $\ge M$ codewords and $d_{min}(C) \ge d$.

Linear Codes

Linear Codes A code $C \subset \{0,1\}^n$ is said to be *linear* if $x,y \in C$ then $x+y \in C \equiv C$ is a vector space in $\{0,1\}^n$.
Generator Matrix *Fact 1* If C is a linear code, then $|C| = 2^k$ for some int k and there exists a $k\times n$ binary matrix G s.t. $C = \{[u_1...u_k]G : (u_1...u_k) \in \{0,1\}^k\} \equiv C$ is a raw space of G.
Parity-check Matrix *Fact 2* If C is a linear code with $|C| = 2^k$, then there exists a $n\times(n-k)$ matrix H s.t. $C = \{\bar{x} : \bar{x}H = [0...0]\}$.
Hamming Codes *Fact.* For the BSC, given $R < C, \varepsilon > 0$ there exists a liner code C s.t. $rate \ge R, p(error) \le \varepsilon$. The *Hamming weight* $w_H(x)$ of a vector $x \in \{0,1\}^n$ is $w_H(x) = \sum_{i=1}^{n} \mathbb{1}\{x_i \neq 0\} = d_H(x, 0...0)$. Given a linear code C, let $w_{min}(C) = min_{x\in C, x\neq0} w_H(x)$.
Theorem For a linear code C, let $w_{min}(C) = d_{min}(C)$.
Field $(\mathbb{F},+,.)$ is a *field*. $a,b \in \mathbb{F} \Rightarrow a+b \in \mathbb{F}, a.b \in \mathbb{F}$. *Properties* $a+b = b+a$, $a.b = b.a$, $(a+b)+c = a+(b+c)$, $(a.b).c = a.(b.c)$, $\exists O \in \mathbb{F}$ s.t. $0+a = a$, $\exists 1 \in \mathbb{F}$ s.t. $1.a = a$, $\forall a \in \mathbb{F}, \exists(-a)$ s.t. $a+(-a) = 0$, $\forall a \neq 0, \exists(a^{-1})$ s.t. $a.(a^{-1} = 1)$, $a.(b+c) = a.b + a.c$. *Theorem "Galois"* Any finite field is isomorphic to one of $\mathbb{F} = \{$ polynomial of degree $< k\}$ with coeff. $+$: poly additive $modp$. : poly mult. $modp$ cond $x^k = ...$ (degree $\le k-1$).
Corollary Finite field exist only with $|\mathbb{F}| = p^k$ (prime power). *Linear codes (general F)* $C \subset \mathbb{F}^k$ is linear if $\forall x,x' \in C, \forall a,a' \in \mathbb{F}, a.x + a'.x' \in C$.
Reed-Solomon Codes Given a field $\mathbb{F}, n \le |\mathbb{F}|, k \le n$. Construct a linear code C with $C \subset \mathbb{F}^n$ and $|C| = |\mathbb{F}|$ as follows : Pick $\alpha_1 \in \mathbb{F}, \alpha_2 \in \mathbb{F}$, ... $\alpha_n \in \mathbb{F}$ s.t. $\alpha_i \neq \alpha_j$ for $i \neq j$. with $u(D) = u_0 + u_1D + ... + u_{k-1}D^{k-1}$, $C = \{(u(\alpha_1),...,u(\alpha_n)) : (u_0...u_{k-1}) \in \mathbb{F}^k\}$ called a *Reed - Salomon code*. *Obs.* If C is a RS code, then C is linear and $d_{min}(C) = w_{min}(C)$. *Theorem* An (n,k) RS code has $d_{min} = n-(k-1)$. Any code $C \subset \mathbb{F}^k$ with $|C| > |\mathbb{F}|^{k-1}$ has $d_{min}(C) \le n-(k-1)$.
Polar Codes *Perfect channel* $p(y|0).p(y|1) = 0$, $C = 1$, trivial code : $m = 0 \to 0$ or $1 \to 1$. *Useless channel* $p(y|0) = p(y|1)$ i.e. Y indpt of X, $C = 0$, trivial code : $m = 0 \to X = 0$ is optimal. *Mediocre channel* If we can convert a mediocre channel into a mixture of the "perfect" and "useless" channels, then communication can take place by means a trivial codes. *Conversion method* $2I(p) = I(p^-) + I(p^+)$, $I(p^-) \le I(p) \le I(p^+)$

Maths

Disjoint $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Credits

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

Rendered January 9, 2016. Written by Lucile Madoulaud.