Information Theory and Coding

February 2012

This summary was kindly written by Marc Zimmerman and Christopher Chiche

1 Source Coding

1.1 Introduction

Singularity A code $\mathcal{C}$ is singular if $\exists u \neq v$ s.t $\mathcal{C}(u) = \mathcal{C}(v)$

Uniquely decodable A code $\mathcal{C}$ is uniquely decodable if $\mathcal{C}^*$ is non-singular.

Prefix-free $\Rightarrow$ uniquely decodable

Instantaneous code A code $\mathcal{C}$ is instantaneous if it is **prefix-free.**

1.2 Optimal Codes

Theorem: 1.1 (Kraft) *A collection* $\{l(u) : u \in \mathcal{U}\}$ *can be the length of a prefix-free code iff* $\sum_{u\in\mathcal{U}} 2^{-l(u)} \leq 1$

Theorem: 1.2 (Kraft) *If* $\mathcal{C}$ *is a uniquely decodable code then* $\sum_{u\in\mathcal{U}} 2^{-l(u)} \leq 1$

Entropy The entropy of a source U is defined as

$$H(U) = \sum_{u\in\mathcal{U}} p(u) \log \frac{1}{p(u)} = E\left[\log \frac{1}{p(U)}\right]$$

Lemma 1.3 $\sum p_i \log \frac{q_i}{p_i} \leq 0$ *where* $q_i = 2^{-l_i}$.

Theorem: 1.4 *For any uniquely decodable code* $\mathcal{C}$ *we have that*

$$E[\mathit{length}(\mathcal{C}(U))] \geq H(U)$$

Theorem: 1.5 *There exists a prefix-free code* $\mathcal{C}$ *with* $E[\mathit{length}(\mathcal{C}(U))] < H(U) + 1$

Theorem: 1.6 (Properties of optimal codes) *1. If $p(u) < p(v)$ then $l(u) \geq l(v)$*

2. Any longest codeword has a "sibling"

3. There is an optimal code s.t. the two least probable symbols are "siblings"

Theorem: 1.7 (Block coding) *For identically and independently distributed (i.i.d) RVs U^n we have*

$$H(U) \leq \frac{1}{n} E[\mathit{length}(\mathcal{C}(U^n))] \leq H(U) + \frac{1}{n}.$$

1.3 Entropy and his Friends

Theorem: 1.8 $0 \leq H(U) \leq \log(|\mathcal{U}|)$ *with equality for the second inequality iff U's are uniformly distributed on $\mathcal{U}$.*

Definition For a random vector $(U_1, \ldots, U_n)$ with distribution $P_{U_1,\ldots,U_n}(U_1, \ldots, U_n)$ we define

$$H(U^n) = E\left[\log \frac{1}{p(U^n)}\right]$$

Theorem: 1.9 *If $\{U_1, \ldots, U_n\}$ are independent RVs we have $H(U^n) = \sum_{i=1}^n H(U_i)$*

Theorem: 1.10 $H(UV) \leq H(U) + H(V)$ *with equality iff U, V are independent*

Theorem: 1.11 $H(X) \geq H(Y)$ *if $Y = f(X)$*

Mutual Information $I(U, V) = H(U) + H(V) - H(UV) = \sum_{u,v} p(u, v) \log \frac{p(uv)}{p(u)p(v)}$. Furthermore, by the preceding theorem, $I(U, V) \geq 0$.

Conditional Entropy $H(U|V) = H(UV) - H(V) = E\left[\log \frac{1}{p(U|V)}\right] = \sum_{v \in V} p(v) H(U|V = v)$

Theorem: 1.12 $I(U, V) = H(U) - H(U|V) = H(V) - H(V|U) \geq 0$ *with equality iff U, V independent.*

Theorem: 1.13 (Chain rule of entropy)

$$\begin{aligned} H(U_1, \ldots, U_n) &= H(U_1) + H(U_2|U_1) + H(U_3|U_1, U_2) + \cdots + H(U_n|U_1, \ldots U_{n-1}) \\ &= \sum_{i=1}^n H(U_i|U^{i-1}) \end{aligned}$$

Corollary 1.14 (Chain rule for conditional entropy)

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z)$$

Theorem: 1.15 (Chain rule of Information)

$$I(U^n;V) = \sum_i I(U_i;V|U^{i-1})$$

Corollary 1.16 (Chain rule for mutual information)

$$I(U^n;V) = H(U^n) - H(U^n|V) = \sum I(U_i;V|U^{i-1})$$

Corollary 1.17 (Chain rule for relative entropy)

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$$

Definition

$$\begin{aligned}
I(U;V|W) &= H(U|W) + H(V|W) - H(UV|W) \\
&= H(UW) + H(VW) - H(UVW) - H(W) \\
&= H(U|W) - H(U|VW) \\
&= H(V|W) - H(V|UW) \\
&= \sum_{u,v,w} p(u,v,w) \log \frac{p(uv|w)}{p(u|w)p(v|w)} \\
&= \sum_w p(w) I(U;V|W=w)
\end{aligned}$$

Theorem: 1.18 *$H(U)$ as a function of the distribution of U is concave.*

Kullback-Leibler Divergence $p(x)$, $q(x)$ two probabilitiy distributions.

$$D(p||q) = \sum p(x) \log \frac{p(x)}{q(x)}$$

1.3.1 Entropy Rate of Stochastic Processes

Definition The entropy rate of a stochastic process $U_1, U_2, \ldots, U_n$ is defined as

$$\lim_{n\to\infty} \frac{1}{n} H(U_1, U_2, \ldots, U_n),$$

when the limit exists.

stationary stochastic process A stochastic process $U_1, U_2, \ldots$ is said to be stationary if the statistics of $U_1 \ldots U_k$ is the same as the statistics of $U_{1+m}, U_{2+m}, \ldots, U_{k+m}$ for every $k \geq 1$, $m \geq 1$.

Theorem: 1.19 *If $U_1, U_2, \ldots, U_n$ is a stationary process, the entropy rate is well defined and*

$$\lim_{n\to\infty} \frac{1}{n} H(U^n) = \lim_{n\to\infty} \frac{1}{n} H(U_n|U^{n-1})$$

Theorem: 1.20 *Given a stationary process with entropy rate H. Then:*

- *for any $R > H$ there is a uniquely decodable code which uses at most R bits/source letter to encode the source.*
- *If $R < H$, no such method exists.*

1.3.2 Asymptotic Equipartition Porperty (AEP)

Definition We define the set of ϵ-typical sequences of length n as

$$T_\epsilon^n = \{u^n | (1-\epsilon)p(a) \leq \frac{1}{n} N(a|u^n) \leq p(a)(1+\epsilon), \forall a \in \mathcal{U}\}$$

Theorem: 1.21 (Properties of T_ϵ^n) *1. $Pr(U^n \in T_\epsilon^n) \to 1$ as $n \to \infty$*

2. *If $u^n \in T_\epsilon^n$ then $2^{-nH(U)(1+\epsilon)} \leq Pr(U^n = u^n) \leq 2^{-nH(U)(1-\epsilon)}$*
3. *$|T_\epsilon^n| \leq 2^{(1+\epsilon)nH}$*
4. *For n large enough $(1-\epsilon)2^{(1-\epsilon)nH} \leq |T_\epsilon^n|$*

Theorem: 1.22 *(Interpretations of Kullback Liebler Divergence)*

- *$Pr(U^n \in T_\epsilon^n) \doteq 2^{-nD(p||q)}$*
- *$E[length[\mathcal{C}(\mathcal{U})]] = D(p||q) + H(p)$ if $\mathcal{U}$ has distrib p but is encoded with distrib q*

Remark $u^n \in T_\epsilon^n(p) \Rightarrow Pr(U^n(q) = u^n) = 2^{-n(D(p||q)+H(p))(1\pm\epsilon)}$

1.4 Universal Source Coding

Coding types memo : 1 - Fixed to variable (Huffman) 2 - Fixed to fixed (coding barel?) 3 - Variable to fixed (Dictionary) 4 - Variable to variable (LZ)

1.4.1 Variable-to-fixed length coding - Tunstall algorithm

Definition A dictionary $\mathcal{D}$ is *valid* if any infinite source sequence has a prefix in $\mathcal{D}$.

Definition A *parser*, given a dictionary $\mathcal{D}$, produces the longest word in the dicitonary which is a prefix of the sequence it is parsing and then repeats.

Definition A dictionary $\mathcal{D}$ is *prefix-free* if no dictionary word is a prefix of another. If a dicitonary is valid and prefix-free

1. Every sequence can be parsed
2. The parser can operate without looking ahead
3. The parsing is unique

Theorem: 1.23 *If a memoryless source $U_1, U_2, \ldots$ is parsed by a valid p.f. dictionary, the entropy $H(W)$ of the parsed word satisfies*

$$H(W) = H(U)E[\text{length}(W)]$$

Theorem: 1.24 (Tunstall algorithm) *1. Start with the root as intermediate node and all level 1 nodes as leaves.*

2. If number of leaves is equal to the desired dictionary size stop.

3. Otherwise, pick the highest probability leaf, make it an intermediate node and grow K leaves on it. Goto step 2

Proposition 1.25 *By choosing b (# of binary digits) large, we are choosing M (# of words) large and thus $E[L]$ large. This makes the term excess of $H(U)$ approach zero. Thus, by taking a large dictionary, the number of bits per source letter the scheme uses can be made as close to $H(U)$ as desired.*

$$\frac{b}{E[L]} < H(U) + \frac{1}{E[L]}[\log(1/P_{min}) + \log(1 + (K-1)/M)].$$

1.4.2 Lempl-Ziv Algorithm

Lemma 1.26 *If a string $u_1, \ldots, u_n$ with $u_i \in \mathcal{U}$, $|\mathcal{U}| = J$ as a concatenation of c* distinct *words $u_1, \ldots, u_n = w_1, \ldots, w_c$, then $n \geq c \log_J \frac{c}{J^3}$*

Compressibility of a string Given an IL encoder E and a string u_1^n.

- $\rho_E(u_1^n) = \frac{1}{n}\text{length}(y_1^n)$
- (E with s states) $\rho_s(u) \lim_{n\to\infty} \sup \rho_s(u_1^n)$
- $\rho(u) = \lim_{s\to\infty} \rho_s(u)$

Theorem: 1.27 *For any IL-encoder with s states*

$$\text{length}(y_1^n) \geq c(u_1^n) \log_2(c(u_1^n)/(8s^2))$$

Theorem: 1.28 *On any infinite string $U_1, U_2, \ldots$ LZ will perform at least as well as any F.S.M.*

Theorem: 1.29 *Suppose $U_1, U_2, \ldots$ is a stationary stochastic process with entropy rate H. Then $E[\rho_{LZ}(u_1^\infty)] \leq H$, with $\rho_{LZ}(u_1^n) =$ # of bits LZ produces when fed $u_1 \ldots u_n$*

2 Data transmission

Theorem: 2.1 *If we have a memoryless channel, used without feedback. Then*

$$P(Y_1 = y_1 ... Y_n = y_n | X_1 = x_1 ... X_n = x_n) = \prod_{i=1}^{n} P(y_i | x_i)$$

Capacity Given a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, $P(Y = y|X = x) = P(y|x)$, we have $C = \max_{p_x} I(X, Y)$

- Capacity of BSC (p switch probability): $C_{BSC} = 1 - H(p)$
- Capacity of Z-Channel: $C_Z = \log_2(1 + (1-p)p^{p/(1-p)})$

Theorem: 2.2 (Fano's Inequality) *If U, V are RVs in the same alphabet $\mathcal{U}$ then :*

$$H(U|V) \leq h_2(p) + p \log_2(|\mathcal{U}| - 1)$$

Where $p = Pr(U \neq V)$ and $h_2(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$

Corollary 2.3 *If $U_1 ... U_L, V_1 ... V_L$ are RVs taking value in the alphabet $\mathcal{U}$, then:*

$$\frac{1}{L} H(U^L | V^L) \leq h_2(\overline{p}) + \overline{p} \log(|\mathcal{U}| - 1)$$

Where $\overline{p} = \frac{1}{L} \sum_{i=1}^{L} Pr(U_i \neq V_i)$

DMC Discrete Memoryless Channel

Theorem: 2.4 *If $X_1 ... X_n$ is the input to a DMC (without feedback) and $Y_1 ... Y_n$ is the output, then:*

$$I(X^n, Y^n) \leq \sum_{i=1}^{n} I(X_i | Y_i) \leq nC$$

Where C in the capacity of the DMC.

Theorem: 2.5 (Data processing Inequality) *If $A - B - C$ forms a Markov Chain, then*

$$I(A; B) \geq I(A, C)$$

Corollary 2.6 *If $A - B - C - D$ is a Markov Chain, then*

$$I(A, D) \leq I(B, C)$$

Block Codes Given a channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ a block code with block length n and M codewords is a mapping: $Enc : \{1...M\} \to \mathcal{X}^n$. A decoder is a mapping: $Dec : \mathcal{Y}^n \to \{?, 1.2...M\}$. The rate of such a code is $R = \frac{1}{n} \log_2 M$ (bits/channel use).

Probability of error for a message M

$$P_{e,m} = P(\{y^n : Dec(y^n) \neq m\}|x^n = Enc(m)) \qquad m = 1, 2, ...M$$

Theorem: 2.7 *Given a DMC with* $C = \max_{p_x} I(X, Y)$*,* $R < C$*,* $\epsilon > 0$*, there exists a block code with rate* $\geq R$*,* $P_{e,max} < \epsilon$

So, for a DMC, the quantity C is a fundamental quantity, namely "at rates up to C we can communicate reliably" [Achievability] and "for rates $> C$ this is not possible" [Converse]

2.1 Communication with feedback

Theorem: 2.8 *For a DMC, feedback does not increase capacity*

3 Convex optimization

Convex function $f : \mathcal{S} \to \mathbb{R}$ is convex if : $\forall u, v \in \mathcal{S}, 0 \leq \lambda \leq 1 : f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v)$

Convex Set A set $\mathcal{S}$ is convex if: $\forall u, v \in \mathcal{S}, 0 \leq \lambda \leq 1 : (\lambda u + (1 - \lambda)v) \in \mathcal{S}$

Concave function A function f is said to be concave if $(-f)$ is convex

Corollary 3.1

1. *If f is convex then* $\forall u_1, ...u_k \in \mathcal{S}, \lambda_1...\lambda_k \geq 0, \sum \lambda_i = 1 : f(\sum_{i=1}^{k} \lambda_i u_i) \leq \sum_{i=1}^{k} f(\lambda_i u_i)$
2. *If $\mathcal{U}$ is a random variable and f is convex, then* $f(E[U]) \leq E[f(U)]$
3. *If f is convex,* $\begin{cases} a \geq 0 \\ a \leq 0 \end{cases}$ *then* $a \cdot f$ *is* $\begin{cases} \textit{convex} \\ \textit{concave} \end{cases}$
4. f_1, f_2 *convex* $\to f_1 + f_2$ *is convex and* $\max(f_1, f_2)$ *is convex*

Theorem: 3.2 *If* $[a, b] \to \mathbb{R}$ *and suppose f is twice differentiable and suppose* $f''(x) \geq 0$ *for* $x \in [a, b]$ *then f is convex*

Theorem: 3.3 *Suppose we are given a DMC* $(\mathcal{X}, \mathcal{Y}, p(y|x))$ *and we set* $f(p) = I(X, Y)$ *when X has distribution p, then f is a concave function of p.*

3.1 Maximizing concave functions over the simplex

Theorem: 3.4 (Kuhn-Tucker conditions for optimality) *A necessary condition for $q \in \mathcal{S}$ to maximize a function f is: $\forall k, j$ s.t $q_j > 0, \frac{\partial f}{\partial q_k} \leq \frac{\partial f}{\partial q_j}$, which is equivalent to: there is some μ such that:*

$$\frac{\partial f}{\partial q_j} = \mu \qquad \forall j \ s.t\ q_j > 0$$
$$\frac{\partial f}{\partial q_k} \leq \mu \qquad \forall k \ s.t\ q_k = 0$$

Theorem: 3.5 *If f: simplex of $K-1$ dimensions $\to \mathbb{R}$, f is concave. then $(p_1..p_K)$ maximizes f if and only if **KT** conditions.*

Theorem: 3.6 *A distribution p_x minimizes $I(X;Y)$ if and only if*

$$\exists \mu \ s.t \sum_y p(y|x) \log \frac{p(y|x)}{p(y)} = \mu \qquad \forall x \ s.t\ p_X(x) > 0$$
$$\leq \mu \qquad \forall x \ s.t\ p_X(x) = 0$$

Furthermore, $\mu = C$

Theorem: 3.7 *The capacity achieving output distribution is unique. Also, the capacity achieving output has $p_Y(y) > 0$ for all y which can be reached for some input x*

Theorem: 3.8 *Suppose $p(x)$ is any input distribution (not necessarily capacity achieving). Then:*

$$\sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \geq C \qquad \textit{for some } x \textit{ with } p(y) \textit{ the output distribution corresponding to } p(x)$$

Corollary 3.9 *For any input distribution $p(x)$:*

$$\sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \leq C \leq \max_x \sum_y p(y|x) \log \frac{p(y|x)}{p(y)}$$

with equality on second inequality if $C = \min_{p(x)} \max_x \sum_y p(y|x) \log \frac{p(y|x)}{p(y)}$

Theorem: 3.10 *For any DMC there is an input distribution $p(x)$ which achieves capacity and has $\text{Support}(p) = \{x : p(x) > 0\})$ of size at most $|\mathcal{Y}|$*

3.2 Communications with cost constraints

$C(\beta) = \max_{p(x_i), E[b(x_i)] \leq \beta} I(X;Y)$

Definition An encoder of rate $\frac{1}{n}$ is said to obey a **max-cost constraint** β if $\frac{1}{n}\sum_{i=1}^{n} b(x_i(m)) \leq \beta$ for every m

Definition An encoder of rate $\frac{1}{n}$ is said to obey an **average-cost constraint** β if $\frac{1}{M}\sum_{m=1}^{M} \frac{1}{n}\sum_{i=1}^{n} b(x_i(m)) \leq \beta$

Theorem: 3.11 *If $R < C(\beta), \epsilon > 0$, then there exists a block Encoder/Decoder such that:*

1. $\frac{1}{n}\log M \geq R$
2. $Pr(\hat{m} \neq m | m \text{ is sent}) < \epsilon \qquad \forall m$
3. $\frac{1}{n}\sum_{i=1}^{n} b(X_i(m)) < \beta + \epsilon \qquad \forall m$

4 Channels with continuous valued input/output

Definition $h(x) \triangleq \int_X p(x) \log \frac{1}{p(x)} \mathrm{d}x = E[\log \frac{1}{p(x)}]$

Theorem: 4.1 (Continuous Stuff) 1. *If $Y = X + cst$ then $h(Y) = h(X)$*

2. *if $Y = aX$ then $h(Y) = h(X) + \log|a|$*
3. *if $X \sim N(\mu, \sigma^2)$ then $h(x) = \frac{1}{2}\log(2\pi e \sigma^2)$*
4. *A constant RV has h equal to $-\infty$*
5. *Differential entropy of jointly Gaussian variables:$X \sim N(\mu, \mathbf{K})$, $h(X^n) = \frac{1}{2}\log\left((2\pi e)^n |\mathbf{K}|\right)$*

Definition It p and q are two probability densities, we define $D(p||q) = \int_X p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$ as the divergence between p and q.

Theorem: 4.2 *$D(p||q) \geq 0$ with $= 0$ if and only if $p = q$*

Corollary 4.3 1. *Suppose X is a $\mathbb{R}$-valued RV which takes values in $[a,b]$. Then $h(X) \leq h(Uniform[a,b]) = \log(b-a)$*

2. *Suppose X is 0-mean, variance σ^2. Then $h(X) \leq \frac{1}{2}\log(2\pi e \sigma^2)$*
3. *Suppose X is μ-mean, $X \geq 0$. Then $h(X) \leq h(Exp(\lambda)) = \frac{1}{\ln 2} - \log_2 \lambda \qquad (\lambda = \frac{1}{\mu})$ ($\Rightarrow$ exp. distribution has worst entropy for non-negative RV.)*

Theorem: 4.4 (Chain-rule) $h(X^n) = \sum_{i=1}^{n} h(X_i | X^{i-1})$

Definition if X, Y are $\mathbb{R}$-valued, $I(X,Y) \triangleq \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y = D(p_{XY}||p_X p_Y)$

Theorem: 4.5 *1. $I(X,Y) \geq 0$, $= 0$ if and only if X and Y are independent.*

2. $h(X|Y) \leq h(X)$, $=$ if and only if X and Y are independent.

3. $I(X^n;Y) = \sum_{i=1}^{n} I(X_i;Y|X^{i-1})$

4. In the differential setting, 1-to-1 transformation preserves $I(\_,\_)$

4.1 Channels with non-discrete alphabet

Definition Given a channel $\mathcal{X}, \mathcal{Y}, p(y|x)$ we say that a rate R is achievable if :

$$\forall \epsilon > 0, \exists Enc/Dec \text{ s.t } rate(Enc) \geq R \text{ and } P(error) < \epsilon$$

Definition The capacity C of a channel is the largest achievable rate $C = \sup\{R|R$ is achievable on the cha

Theorem: 4.6 *For a memoryless channel $\mathcal{X}, \mathcal{Y}, p(y|x)$, $C = \sup_{p(x)} I(X.Y)$*

Theorem: 4.7 $C(\beta) = \sup_{p_x, E[b(x)] \leq \beta} I(X;Y)$

Gaussian channel with power constraint $C = \frac{1}{2}\log(1 + \frac{p}{\sigma^2})$

5 Rudiments of Coding theory

$(7,4)$ **Hamming Code** Binary code with block length 7 where $\vec{x} \in \mathbb{F}_2^n$ is a codeword if and only if it satisfies :

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_7 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

This code has 2^5 codewords. Single error correcting.

Hamming codes $(2^m - 1, 2^m - m - 1)$: # of codewords $= 2^{dimensions}$. Capable of correcting "1 flip"

Remark The Hamming distance d_h is a metric

Theorem: 5.1 *If $d = d_{min}(\mathcal{C})$ then the code $\mathcal{C}$ can correct $\lfloor \frac{d-1}{2} \rfloor$ flips*

Theorem: 5.2 *$d_{min}(\mathcal{C})$ is a good "figure of the merit" in measuring how good a code is.*

Hamming weight of a sequence $\bar{x}$ $w_H(\bar{x}) = \#\{i, x_i \neq 0\}$

Theorem: 5.3 *If $\mathcal{C}$ is a linear code, then $d_{min}(\mathcal{C}) = \min_{\bar{x} \in \mathcal{C}, x \neq 0} w_H(\bar{x})$*

Theorem: 5.4 (Sphere packing bound) *Suppose a codeword $\mathcal{C} \in \mathbb{F}_2^n$ can correct all possible e or fewer flips. Then:*

$$|\mathcal{C}| \sum_{i=0}^{e} \binom{n}{i} \leq 2^n$$

With equality if perfect code

Theorem: 5.5 (Gilbert Varshamov bound) *Given a block length n and d. there is a code $\mathcal{C} \in \mathbb{F}_2^n$ with : $d_{min}(\mathcal{C}) \geq d$, $|\mathcal{C}| \sum_{i=0}^{d-1} \binom{n}{i} \geq 2^n$*

Corollary 5.6 (Gilbert Varshamov for linear codes) *Given n, d there is a linear code $\mathcal{C}$ with $|\mathcal{C}| \sum_{i=0}^{d-1} \binom{n}{i} \geq 2^n$*

Theorem: 5.7 (Singleton bound) *If a code $\mathcal{C} \in \mathbb{F}_2^n$ has $|\mathcal{C}| > 2^k$ (k integer) then $d_{min}(\mathcal{C}) \leq n - k$*

5.1 Reed-Salomon codes

Block codes over alphabets which are algebric fields $\mathcal{C} \in \mathbb{F}^n$.

They are described as follows: Pick $\alpha_1, \alpha_2 ... \alpha_n$ distinct elements of $\mathbb{F}$. The code is going to have $|\mathbb{F}|^k$ codewords, to each$\vec{u} \in \mathbb{F}^k$ we will associate a codeword $\vec{x}(\vec{u})$. The rule that describes $\vec{u} \mapsto \vec{x}(\vec{u})$ is as follows:

Given $\vec{u} = (u_0, ... u_{k-1})$, first construct the polynomial $U(D) = u_0 + u_1 D + u_2 D^2 + ... + u_{k-1} D^{k-1}$. Set $\vec{x}(\vec{u}) = (u(\alpha_1), u(\alpha_2), ..., u(\alpha_n))$

Theorem: 5.8 $d_{min}(RS\ code\ \mathcal{C}) = n - k - 1$

- A (n, k) Reed-Salomon code can correct $(n - k)$ erasures
- It is possible to achieve the capacity of the Binary Symetric Channel by using Linear Codes.

6 Rappels

Theorem: 6.1 (Chebychev's inequality) *In the case of a random variable S_n that is the sum of n i.i.d. random variables $X_1, X_2, ..., X_n$ we have :*

$$Pr(|S_n - n\mu| \geq a) \leq \frac{n\sigma^2}{a^2}$$

Theorem: 6.2 (Strong law of large numbers)

$$\forall \epsilon, Pr(\lim_{n \to \infty} |\bar{X}_n - \mu| > \epsilon) = 0 \text{ with } \bar{X}_n = \frac{1}{n}(X_1 ... X_n)$$

Theorem: 6.3 (About markov Chains) *If $X - Y - Z$ is a Markov chain :*

- $I(X, Z|Y) = 0$
- $I(X, Y) \geq I(X, Z)$

Theorem: 6.4 (Jensen's inequality) *If f is a convex function, then $E[f(x)] \geq f(E[x])$*

Sums $\displaystyle\sum_{n=0}^{N-1} r^n = \frac{1 - r^N}{1 - r}$ $\qquad$ $\displaystyle\sum_{n=0}^{\infty} nr^n = \frac{r}{(1 - r)^2}$

K-ary tree If a K-ary tree has n nodes, then it has $1 + (K - 1).n$ leaves.