Definition of a d.r.v

- $\vec{X}$: finite set X of possible outcomes // probability distribution over X

Def of a uniform probability dist

each outcome is equally likely $\implies \Pr(\mathbf{X}=x)=\frac{1}{|X|}$

Joint pd of two drvs

Given $\mathbf{X}$, $\mathbf{Y}$, joint pd has
- drv $(\mathbf{X},\mathbf{Y})$
- set $X \times Y$
- pd $\Pr(\mathbf{X}=x,\mathbf{Y}=y)$ (typically write $\Pr(x,y)$)
- (***independent*** $\impliedby \Pr(x,y)=\Pr(x)\Pr(y)$)

Def. of a conditional distribution

$$(\mathbf{Y}|\mathbf{X}): \qquad \Pr(y|x)=\frac{\Pr(\mathbf{X}=x,\mathbf{Y}=y)}{\Pr(\mathbf{X}=x)}$$

State and use Bayes' thm

$$\Pr(x|y)=\frac{\Pr(y|x)\Pr(x)}{\Pr(y)}$$

Definition of a finite field

Field: set $\mathbb{F}$ plus $+,\times$;

$(\mathbb{F},+)$ abelian, identity 0

$(\mathbb{F}\setminus\{0\},\times)$ abelian, identity 1

distributive $\times/+$ and $+/\times$

$\mathbb{F}$ finite $\implies$ **finite** field

For each prime power q there is a unique finite field order q

(up to isomorphism)

Recall and use basic properties of finite fields

$q=p^n$
- $(GF(q)*,\times)$ is cyclic
- d divides $n \implies$ unique subfield order p^d
- no other subfields
- field has char p
- group A of automorphisms of field is cyclic, $|A|=n$, $a\to a^p$ (***Frobenius automorphism***)

Construct $GF(q^n)$ using an irreducible polynomial over $GF(q)$

Use $x^{k+1}+a_kx^k\ldots+a_0=0 \implies x^{k+1}=-a_kx^k\ldots-a_0$; sub in as necessary

Perform polynomial interpolation (in $GF(q)[x]$)

Set of simultaneous equations

(Or Lagrange interp formula)

Defn of Shannon entropy of a drv

$$\mathrm{H}(\mathbf{X}) = -\sum_{i=1}^{n} p_i \log p_i$$

Compute the Shannon entropy of a drv

per formula

Fundamental Lemma

Given $\sum_i p_i = \sum_i q_i = 1$ (two pds):

$$-\sum_i p_i \log p_i \leq -\sum_i p_i \log q_i;$$

equality $\iff p_i = q_i$ for all i

Joint entropy of two drvs $\leq$ sum of their entropies; prove via Fundamental Lemma

$$\mathrm{H}(\mathbf{X}, \mathbf{Y}) \leq \mathrm{H}(\mathbf{X}) + \mathrm{H}(\mathbf{Y})$$

---

def of

$\mathrm{H}(\mathbf{X}|\mathbf{Y} = y)$, $\mathrm{H}(\mathbf{X}|\mathbf{Y})$

$$\mathrm{H}(\mathbf{X}|\mathbf{Y} = y) = -\sum_x \Pr(x|y) \log \Pr(x|y)$$

"uncertainty in the outcome of $\mathbf{X}$ once we know that the outcome of $\mathbf{Y}$ is y"

Compute

$\mathrm{H}(\mathbf{X}|\mathbf{Y} = y)$, $\mathrm{H}(\mathbf{X}|\mathbf{Y})$, **given X, Y**

$$\mathrm{H}(\mathbf{X}|\mathbf{Y}) = \sum_y \Pr(y) \mathrm{H}(\mathbf{X}|\mathbf{Y} = y)$$

"average amount of uncertainty about the outcome of $\mathbf{X}$ remaining once outcome of $\mathbf{Y}$ is known"

State and prove

$\mathrm{H}(\mathbf{X}, \mathbf{Y}) = \mathrm{H}(\mathbf{Y}) + \mathrm{H}(\mathbf{X}|\mathbf{Y})$ expand out definition

---

State and prove

$\mathrm{H}(\mathbf{X}|\mathbf{Y}) \leq \mathrm{H}(\mathbf{X})$ equality $\iff$ $\mathbf{X}, \mathbf{Y}$ independent

Define $\mathrm{I}(\mathbf{X}|\mathbf{Y})$ and compute it

$\mathrm{I}(\mathbf{X}|\mathbf{Y}) = \mathrm{H}(\mathbf{X}) - \mathrm{H}(\mathbf{X}|\mathbf{Y})$

"reduction in uncertainty associated with $\mathbf{X}$ once we know the value of $\mathbf{Y}$"

Show that $\mathbf{H}(X|Y) \geq 0$

equality $\iff$ $\mathbf{X}, \mathbf{Y}$ independent

Follows from prev. unit

---

show that $\mathrm{I}(\mathbf{X}|\mathbf{Y}) = \mathrm{I}(\mathbf{Y}|\mathbf{X})$

hence "mutual" information

Defn of a discrete memoryless source

- Finite source alphabet, symbols called words
- Sequence $\mathbf{W}_0, \ldots, \mathbf{W}_i$
- $\mathbf{Pr}(\mathbf{W}_j = w_i) = p_i$

$\implies$ the $\mathbf{W}_i$ are independent, identically distributed drvs - entropy of source: $\mathrm{H}(\mathbf{W})$

Defns of

instantaneous	no encoded word is a prefix of any other encoded word
uniquely decipherable	for any sequence S, at most one source message can be encoded as S
compact encoding	u.d. with smallest possible expected encoded word length

---

Kraft's inequality; McMillan's inequality

Kraft: (existence of instantaneous encoding)	$\sum_{i=1}^{m} D^{-n_i} \leq 1$
McMillan: (uniquely decipherable)	$\sum_{i=1}^{m} D^{-n_i} \leq 1$

Note identical!

Every instantaneous code is uniquely decipherable, i.e. [TODO!]

Shannon's noiseless coding theorem

$\mathbf{W}$ a discrete memoryless source

1.
$$\bar{n} \geq \frac{\mathrm{H}(\mathbf{W})}{\log D}$$

2. There exists u.d. encoding with
$$\frac{\mathrm{H}(\mathbf{W})}{\log D} + 1 \geq \bar{n}$$

Perform Huffman coding

- Sort source words by probability
- Put as leaves of tree; build tree by merging least probable nodes

---

Huffman coding produces compact instantaneous encodings

(not unique) (prove by induction: base case 2 words)

Defs of ideal observer decoding; max. likelihood decoding

Given r_j, decode as t_i s.t.

Ideal observer:	max. $Pr(t_i \mid r_j)$
Max. likelihood:	max. $Pr(r_j \mid t_i)$

Ideal observer requires *a priori* message probs; max likelihood does not

---

Def of binary symmetric channel

input, output alphabets both $\{0, 1\}$

flip probability $p < 0.5$

Calculate Hamming dists

Q a finite set; list of elements of Q a codeword;

$\mathcal{C}$ a set of codewords a code;

codewords all same length $\implies$ block code; $Q = \{0, 1\}$ binary code

Hamming distance $d(\mathbf{w}, \mathbf{u}) = |\{i \in \{1, 2, \ldots, n\} | w_i \neq u_i\}$

Prove properties of Hamming dists

e.g.

1. $d(\mathbf{u}, \mathbf{v}) \geq 0$, equality $\iff$ $\mathbf{u} = \mathbf{v}$
2. $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
3. $d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) \geq d(\mathbf{u}, \mathbf{w})$ ***triangle inequality***

In binary sym. channel, NN decoding is equivalent to max. likelihood

NN = choose t_i minimising $d(r_j, t_i)$

(recall $p < \frac{1}{2}$ by definition)

Min. dist of a block code

Largest d s.t. for any $\mathbf{u}, \mathbf{v} \in \mathcal{C}$, $d(\mathbf{u}, \mathbf{v}) \geq d$

$\implies (n, M, d)$-code

Connection between min. dist of a code & error-correcting properties

Max. errors corrected by NN decoding of block code $\mathcal{C}$ with min. dist d:

$$\left\lfloor \frac{d-1}{2} \right\rfloor$$

(see also Ex 1.5)

Capacity of a noisy channel

capacity: $\sup_{\mathbf{R}} \mathrm{I}(\mathbf{T}|\mathbf{R})$

"the greatest possible amount of information that the channel output gives about input to the channel"

Compute capacity of binary sym. channel

$$1 + p \log p + (1-p) \log (1-p)$$

(use formula, max. per derivative)

nth extension of channel with cap. C has cap. nC

Shannon's Noisy Coding Thm

rate R of a binary code of length n with M codewords: $\frac{1}{n} \log_2 M$

binary sym. channel with $0 < R < C$ capacity

$\epsilon > 0$, sequence of integers $M_0, M_1, M_2, \ldots$ with $1 \leq M_i \leq 2^{Ri}$:

there is some integer N_0 and $\mathcal{C}_0, \mathcal{C}_1, \ldots$ s.t. $\mathcal{C}_i$ has length i, M_i codewords, max. error prob $\leq \epsilon$

Basically: if you make your codes big enough, you can make the error as small as you want

Proof (sketch): TODO

$$\boxed{\text{msg}} \xrightarrow{\mathbf{m}G} \text{codeword} \xrightarrow{\text{noise}} \text{rec'd word} \xrightarrow{\text{NN}} \text{codeword'} \rightarrow \boxed{\text{msg'}}$$

Def. of an $[n, k, d]$ code

Square brackets [] $\implies$ linear

$\implies$ vector space over alphabet Q ($\implies$ Q is a ***field***)

dimension of vector space: k n length of code // d min. dist

Use vector space properties to prove simple results

linear code: all linear combos of any words are also words

Def. of linear code as gen. mat + par mat

Gen mat: rows form basis

Gen $\rightarrow$ par mat

Par mat has rank $n - k$; rows are orthogonal to all codewords

Systematic form $(\mathbb{I}_k|A)$ has par mat $(-A^T|\mathbb{I}_{n-k})$

More generally require orthogonality

Deduce dimension, min. dist., etc. from gen/par mats

Every vector space includes $\mathbf{0}$. Hamming weight of word = dist from 0. min weight of code = smallest Hamming weight = min dist

min dist = min # lin. ind. cols in parmat $\implies$ 1 $\iff$ all-zero col; 2 $\iff$ 1 col is multiple of another (equal in binary) ...

Syndrome decoding

- any word $\mathbf{c}$ in code has $H\mathbf{c}^T = 0$ for par mat H (def)

- $\mathbf{r} = \mathbf{c} + \mathbf{e}$ received, $H\mathbf{r}^T = H\mathbf{e}^T$ $(1 \times (n-k))$ is ***syndrome*** of $\mathbf{r}$

- divide vector space into cosets by syndrome (1 codeword per coset) // sort cosets by Hamming weight (=coset leader min weight)

- decode by matching syndrome with coset leader

= fast implementation of NN decoding (with precomputation phase)

☆ : Observation: $\mathbf{e}$ has weight (0 or) 1, then the syndrome of r = HeT is just a scalar multiple of a column of H, say col j: flip jth bit (also applies to non-binary...)

(NB: if we /can/ guaranteed decode m errors then there must be a unique coset leader $\leftarrow$ multiple coset leaders give different results)

Dual code

Take par mat and use as gen mat

Dual of an $[n, k]$ code is $[n, n-k]$ code. ***Orthogonal complement***: dual code is orthogonal to code

---

Find dual codes

Gen $\leftrightarrow$ par

Sphere-packing bound

$$A_q(n,d) \le \frac{q^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i}(q-1)^i}$$

What is it? $A_q(n,d)$ is largest possible M with q-ary (n,M,d) code, i.e. largest possible number of codewords in an (n,d) code.

Def of perfect code

Sphere-packing bound met with equality

$(\implies d$ odd$)$

Hamming codes; binary Hamming codes are perfect

Par mat is all non-zero binary vectors of length k.

Capable of correcting single error

Can be generalised to q-ary Hamming codes; still perfect

$$\mathcal{H}_{2,r} \text{ is a } [\underbrace{2^r-1}_{n}, n-r, 3] \text{ code}$$

$$\text{divide out scalar multiples} \mathcal{H}_{q,r} \text{ is a } [\underbrace{\frac{q^r-1}{q-1}}_{n}, n-r, 3] \text{ code}$$

Singleton bound

$$A_q(n,d) \le q^{n-d+1}$$

equality $\implies$ maximum-distance separable = ***MDS*** code

---

RS codes

Take n elements of a q-ary alphabet, $q \ge n$. Take polynomials deg $\le k$ for some $k \le n$; each codeword is the result of evaluating a polynomial at the n elements

---

RS codes are MDS codes

$[n, k, n-k+1]$ codes. Show that

(1) linear [linear combo of codewords is also a codeword]

(2) $d \ge n-k+1$ (by polynomial interp.)

(3) singleton bound $d \le n-k+1 \implies d = n-k+1$

(in fact, they are only non-trivial MDS codes known) this is not quite true as e.g. 2.13 is not RS!

✓

efficient decoding

✗ need big alpha

<u>Bounds</u> (U6 & U7):

<u>Bounds on code size</u>:

Sphere-packing: $A_q \leq \dots$ *perfect* codes meet : e.g. Hamming codes

Gilbert-Varshamov: $A_q \geq \dots$

Singleton: $A_q \leq \dots$ *MDS* codes: e.g. R-S codes

$\rightsquigarrow$ Asymptotic singleton $\alpha_q \left(\frac{n}{d} \rightarrow\right)$

<u>Bound on code length</u>:

Griesmer : simplex codes

A linear code C has minimum Hamming distance d if and only if its parity check matrix H has a set of d linearly dependent columns but no set of $d-1$ linearly dependent columns.

Reminder: 2 columns are l.i. in binary field $\iff$ they are identical (nothing so simple for 3)

Unit 6 Learning Outcomes

State + prove Gilbert-Varshamov bound

$$A_q(n,d) \geq \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i}$$

Proof: count elements distance max $d-1$ from a given codeword

---

Griesmer bound

$$n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil$$

Proof based on residual codes

Puncturing a code

Delete certain coordinates and then collapse code to unique codewords

---

Find punctured code

support of a codeword $\mathbf{c}$ is its nonzero coordinates

residual code wrt $\mathbf{c}$: puncture code on support of $\mathbf{c}$

don't forget to remove duplicates

result is linear code

Describe simplex codes

duals of Hamming codes. Take r-dimensional simplex code over $GF(q)$:

all non-zero codewords have same weight q^{r-1}

satisfies Griesmer bound with equality

Authentication, data integrity, confidentiality for data security

auth: who am I really talking to

integrity: msg didn't get corrupted

confidentiality: eavesdropper can't learn msg

Symmetric encryption

Common key k

Encryption algo Enc s.t. $c = Enc(k,m)$ // Matching Dec

Require $\text{Dec}(k, \text{Enc}(k,m)) = m$

Kerckhoff's principle

Assume eavesdropper knows everything except k: key space, message space, ciphertext space, Enc, Dec.

Better than "security through obscurity" principle

Caesar cipher: why insecure

addition mod 26:

small key space ; many message properties preserved

Substitution cipher: why insecure

frequency analysis!

Unit 7 Learning Outcomes

Describe perfect secrecy

for all $m \in M, c \in C$, $\Pr(m|c) = \Pr(m)$

"knowing the ciphertext does not help eavesdropper guess message"

Prove one–time pad provides perfect secrecy

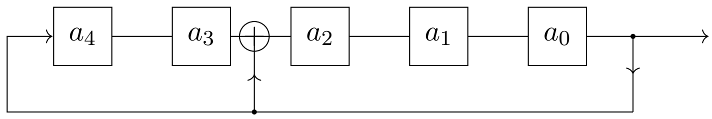
$|M| = |C| = |K|$

Symm. scheme with perfect secrecy requires $|K| \geq |M|$

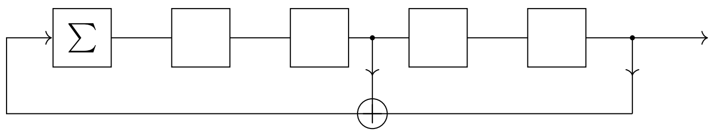
Means scheme is expensive and often not practical

Defns of Galois / Fib LFSRs; understand diags

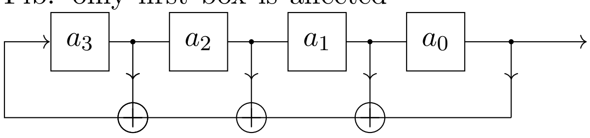
Stream ciphers.



Galois: xor *en route*



Fib: only first box is affected



$$a_4 = a_3 + a_2 + a_1 + a_0 \implies P : x^4 + x^3 + x^2 + x + 1$$

Prove that r-bit binary LFSR has max output seq. $2^r - 1$

If it gets to 0 it gets stuck

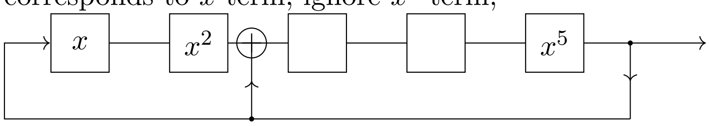
---

m-sequence

output sequence of length $2^r - 1$ (m for maximum)

Use primitive poly over $GF(2)$ to construct Galois LFSR

1st & last are always 1; rightmost bit is always fed back, corresponds to last bit, leftmost bit corresponds to x term, ignore x^0 term;



$1 + x^2 + x^5$

TODO check!

NB Fib LFSR by recurrence relation // set up from poly

Set of possible sequences output by max period r-bit LFSR is vector space dim. r

and therefore a linear code

Unit 8 Learning Outcomes

Definition of a $(k; n)$ threshold scheme

n players

Any k or more players can recover secret

No $k-1$ or fewer players learns anything at all (ie all possibilities EQUALLY LIKELY)

Shamir scheme

$q = p^r; q > n; s \in GF(q)$

Pick $f(x)$ polynomial degree $k-1$ (or less) with coefficients in $GF(q)$ / constant term s

Recover secret by polynomial interp.

Linear secret-sharing scheme

Vandermonde matrix: geometric progression. Property: any k rows are lin. independent

$M : (n+1) \times k$: contains powers of elements of $GF(q)$

$\mathbf{r} = (s\ r_1\ \ldots\ r_{k-1})^T$ (r_i chosen at random); $M\mathbf{r}$ is secret + shares

Given any k players, their combo is lin. independent; can recover secret;

$k-1$ players + target: likewise lin. independent $\rightarrow$ can't recover secret

NB: Any scheme that can be constructed from matrix is a ***linear*** secret-sharing scheme; efficient to describe, build, and use

Shamir vs linear: shares are same! (VdM as polynomials...)

Link between Shamir and RS codes

Set of potential vectors is the codewords of an RS code

$[n+1, k, n+2-k]$ RS code

Can use any MDS code over $GF(q), q > n$. Rows are ***distribution rules***; code (matrix) is published in advance but only dealer knows which row is distributed this time

---

Information rate of a secret-sharing scheme

$$\min_i \frac{\log_2(|K|)}{\log_2(|S_i|)}$$

S_i is set of possible shares for player i

Shamir: rate 1

A perfect secret-sharing scheme has information rate ≤ 1

Size of share space for each player must be at least as big as the secret space

Rate = 1: scheme is ***ideal***

(t_1, t_2, n) **ramp scheme**

up to t_1: no information

t_2 or more: full information

$t_1 \rightarrow t_2$: maybe some information ¯_('')_/¯

Construct ramp scheme from error-correcting code

Words of code are distribution rules; last s coordinates are secret

Define c-TA, c-FP codes

Want to track piracy;

code // NN-decoding

Assume pirated content is generated by coalitions of size c

c-TA: can always find one of the pirates c-FP: **weaker**: can't necessarily find pirate, but can't frame anyone else ("frameproof")

Construct them from error-correcting codes

Any q-ary, length n code min. dist. d, $d > n - \left\lceil \frac{n}{c^2} \right\rceil$ is a c-TA code ($c \geq 2$)

(e.g. a RS code)

Prove that every c-TA code is a c-FP code

"if a set S of up to c pirates could frame some user $\mathbf{y} \in \mathcal{C}$, $\mathbf{y}$ is its own NN $\implies$ lies in S: S cannot frame any users whose words are not in S"

Interested in c-FP codes that are larger than the largest c-TA codes (else why bother), e.g. $[n, \left\lceil \frac{n}{c} \right\rceil, n - \left\lceil \frac{n}{c} \right\rceil + 1]$ RS code is a c-FP code ($c \geq 2$): $q^{\left\lceil \frac{n}{c} \right\rceil}$ codewords $\geq q^{\left\lceil \frac{n}{c^2} \right\rceil}$

Prove whether given codes satisfy definitions to be c-TA, c-FP

Theorems etc

Thm U1#1.8. (Bayes)

$$\Pr(x|y) = \frac{\Pr(y|x)\Pr(x)}{\Pr(y)}$$

Thm U1#2.5. For every prime power $q = p^n$ there exists a unique field $GF(q)$ with q elements, with:

- $(GF(q)^*, \times)$ is a cyclic group (thus there is α "primitive element" of $GF(q)^*$)
- d divides $n \implies GF(q)$ has unique subfield order p^d; these are the *only* subfields of $GF(q)$ ($GF(p) \cong \mathbb{Z}_p$
- $a \in GF(q) : pa = 0$ ("characteristic p")
- group of automorphisms of $GF(q)$ is cyclic order n, generated by $a \to a^p$ ("Frobenius automorphism")

Ex U1#2.3. $\mathbb{Z}_n$ is a field $\iff$ n is prime

Ex U1#2.6. no. of prim. elts of $GF(q = p^m)$ is number of things coprime to $q-1$; no. of prim polys is $\frac{\text{prim elts}}{m}$

Ex U1#2.7. Char $p \implies (a+b)^p = a^p + b^p$

Ex U1#2.8. $GF(q^n)$ is a vector space over $GF(q)$ (any prime power q)

Page U1#7.

- Unique polynomial factorisations

Thm U2#1.1. (Polynomial interp.) $(x_i, y_i) \in GF(q = p^n)^2$, for $i = 0, 1, \ldots, n$; no duplicate x_is: $\implies$ there is a unique polynomial $f \in GF(q)[x]$ with $y_i = f(x_i)$ for all i

Ex U2#1.3. Lagrange interp.

$$f(x) = \sum_{i=0}^{n} y_i f_i(x)$$

where

$$f_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}$$

Thm U2#2.7. Shannon entropy ($\mathrm{H}(\mathbf{X}) = -\sum_i p_i \log p_i$)

- $\mathrm{H}(\mathbf{X})$ is a continuous function of the probabilities
- If $\mathbf{Y}_n$ is uniform rv with n outcomes, $H(\mathbf{Y}_{n+1}) > \mathrm{H}(\mathbf{Y})$
- $\mathbf{Z}$ with two possible outcomes;

$$\mathrm{H}(\mathbf{Z}, \mathbf{X}) = \mathrm{H}(\mathbf{Z}) - \Pr(z_1) \sum_i \Pr(x_i|z_1) \log \Pr(x_i|z_1) - \Pr(z_2) \sum_i \Pr(x_i|z_2) \log \Pr(x_i|z_2)$$

- $\mathrm{H}(\mathbf{X}) \geq 0$; equality $\iff$ only one possible outcome
- $\mathrm{H}(\mathbf{X}) \leq \log n$; equality $\impliedby$ $\mathbf{X}$ has uniform dist
- any function satisfying these properties is (constant multiple of) Shannon entropy!

Lemma U2#2.11. (Fundamental Lemma) $\sum_i p_i = \sum_i q_i = 1$ (p_i, q_i positive real numbers):

$$-\sum_i p_i \log p_i \leq -\sum_i p_i \log q_i$$

Proof using $\ln x \leq x - 1$

Thm U2#2.12.

$$\mathrm{H}(\mathbf{X}, \mathbf{Y}) \leq \mathrm{H}(\mathbf{X}) + \mathrm{H}(\mathbf{Y}),$$

equality $\iff$ $\mathbf{X}, \mathbf{Y}$ independent

Ex U2#2.18. $\mathrm{H}(\mathbf{X}|\mathbf{X}) = 0$

Ex U2#2.19. $\mathbf{X}, \mathbf{Y}$ independent $\implies$ $\mathrm{H}(\mathbf{X}|\mathbf{Y}) = \mathrm{H}(\mathbf{X})$

Thm U2#2.20.

$$\mathrm{H}(\mathbf{X}, \mathbf{Y}) = \mathrm{H}(\mathbf{Y}) + \mathrm{H}(\mathbf{X}|\mathbf{Y})$$

Cor. U2#2.21. $\mathrm{H}(\mathbf{X}|\mathbf{Y}) \leq \mathrm{H}(\mathbf{X})$, equality $\iff$ $\mathbf{X}, \mathbf{Y}$ independent

---

Thm U3#1.2. $\mathrm{I}(\mathbf{X}|\mathbf{Y}) \geq 0$, equality $\iff$ $\mathbf{X}, \mathbf{Y}$ independent

---

Ex U5#3.8.

$$\mathrm{H}(\mathbf{U}|\mathbf{V}) \leq \mathrm{H}(\mathbf{U}|\mathbf{V}, \mathbf{W}) + \mathrm{H}(\mathbf{W})$$

Thm U3#1.3. $\mathrm{I}(\mathbf{X}|\mathbf{Y}) = \mathrm{I}(\mathbf{Y}|\mathbf{X})$

Thm U3#2.12. **(Kraft)** Alphabet $|\Sigma| = D$ Instantaneous encoding with word lengths $n_i \iff$
$$\sum_{i=1}^{m} D^{-n_i} \leq 1$$

Thm U3#2.13. **McMillan** Uniquely decipherable encoding $\iff$
$$\sum_{i=1}^{m} D^{-n_i} \leq 1$$
Prove "if Kraft then exists" and "if exists then McMillan" and other directions follow by "if instantaneous then u.d."

Thm U3#2.15. **(Shannon's Noiseless Coding)** $\mathbf{W}$ a discrete memoryless source with alphabet W of w_i with probs p_i, entropy $\mathrm{H}(\mathbf{W}) = -\sum_i p_i \log p_i$: For any uniquely decipherable encoding of W over alphabet Σ, $|\Sigma| = D$, into codewords of lengths n_i:
$$\frac{\mathrm{H}(\mathbf{W})}{\log D} \underbrace{\leq}_{\text{Any u.d}} \bar{n} \underbrace{\leq}_{\text{u.d. must exist}} \frac{\mathrm{H}(\mathbf{W})}{\log D} + 1$$

Page U4#4. Huffman coding is a compact encoding

Ex U4#2.11. Hamming distance properties:
1. $d(\mathbf{u}, \mathbf{v}) \geq 0$, equality $\iff \mathbf{u} = \mathbf{v}$
2. $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
3. $d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) \geq d(\mathbf{u}, \mathbf{w})$ ***triangle inequality***

Thm U5#1.1. For the binary symmetric channel, NN decoding is equivalent to max. likelihood decoding

---

Thm U5#1.4. NN decoding of a block code with min. dist d can correct up to $\left\lfloor \frac{d-1}{2} \right\rfloor$ errors

---

Thm U5#2.2. Capacity of binary symmetric channel with error prob p:
$$1 + p \log p + (1-p) \log (1-p)$$

Page U5#5. If channel has cap. C, its nth extension has cap nC

Thm U5#3.4. **(Shannon's Noisy Coding Thm)** For a binary symm. channel with cap C and rate $0 < R < C$. Given $\epsilon > 0$, sequence of integers $M_0, M_1, M_2, \ldots$ with $1 \leq M_i \leq 2^{Ri}$:
there is some integer N_0 and sequence $\mathcal{C}_0, \mathcal{C}_1, \ldots$ s.t. $\mathcal{C}_i$ has length i, M_i codewords, max. error prob $\leq \epsilon$ for all $i \geq N_0$
Basically: if you make your codes big enough, you can make the error as small as you want

---

Thm U5#3.5. (Chebyshev) For any real $a > 0$

$$Pr(|\mathbf{X} - E(\mathbf{X})| \geq a) \leq \frac{\text{var}(\mathbf{X})}{a^2}$$

Lemma U5#3.6. $0 \leq p \leq \frac{1}{2}$:

$$\sum_{r=0}^{\lfloor pn \rfloor} \binom{n}{r} \leq 2^{nh(p)}$$

(where $h(p) = -p \log p - (1-p) \log{(1-p)}$) To prove, assume pn integer; write $1 = (p + (1-p))^n$; do magic

---

Thm U5#3.7. C capacity of discrete memoryless channel. $R > C$: no sequence of codes $\mathcal{C}^i$ with $\mathcal{C}^i$ having length i and 2^{nR} codewords with error probability tending to 0 as $n \to \infty$

Lemma U5#3.9. (Fano) $\mathbf{X}, \mathbf{Y}$ drvs with input set = output set $X = Y$; let $\mathbf{Z}$:

$$\mathbf{Z} = \begin{cases} 0 & \mathbf{X} = \mathbf{Y} \\ 1 & \mathbf{X} \neq \mathbf{Y} \qquad \approx \text{(decoding error)} \end{cases}$$

Then:

$$\mathrm{H}(\mathbf{X}|\mathbf{Y}) \leq \mathrm{H}(\mathbf{Z}) + \Pr(\mathbf{Z} = 1) \log{(|X| - 1)}$$

Ex U6#1.5. Min. weight of a linear code is its min. dist

Ex U6#1.6. $\mathcal{C}$ a linear code, G its gen mat: elementary row ops, permuting columns, multiplying columns by nonzero scalars $\implies \mathcal{C}'$ equivalent to $\mathcal{C}$

Ex U6#1.7. Every $[n, k, d]$ code is equivalent [but **not equal!**] to a code with gen mat in form $(\mathbb{I}_k|A)$

Ex U6#1.10. H is par mat $\implies$ codewords are all $\mathbf{c}$ s.t. $H\mathbf{c}^t = 0$

Ex U6#1.11. If gen mat is $(\mathbb{I}_k|A)$ then par mat is $(-A^T|\mathbb{I}_{n-k})$

Ex U6#1.14. G gen mat for code is par mat for dual code

Ex U6#1.15. Dual of an $[n, k]$ code is an $[n, n-k]$ code

Thm U6#2.1. (Sphere-packing)

$$A_q(n, d) \leq \frac{q^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i}(q-1)^i}$$

Ex U6#2.2. If q-ary (n, M, d) code is perfect, d is odd

Page U6#8. Any 2 columns lin. ind $\implies$ no word has weight ≤ 2

Thm U6#2.5. Binary Hamming codes are perfect

Ex U6#2.6. 🏳️‍🌈 non-binary Hamming codes too

Thm U6#2.10. Singleton Bound

$$A_q(n, d) \leq q^{n-d+1}$$

Cor. U6#2.11. For $\mathcal{C}$ an $[n, k, d]$ code over $GF(q)$:
$\dim \mathcal{C} \leq n - d + 1$

Thm U6#2.16. RS codes are $[n, k, n-k+1]$ codes, i.e. MDS codes

---

Thm U7#1.1. (**Gilbert-Varshamov**)

$$A_q(n, d) \geq \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i}$$

Thm U7#2.1. (**Asymptotic singleton bound**) define $\alpha_q(\delta)$ as asymptotic limit of A_q, with δ the limit of the relative distance $\frac{d}{n}$. Then

$$\alpha_q(\delta) \leq 1 - \delta$$

Lemma U7#3.2. Residual code obtained by puncturing on the support of some codeword weight w is an $[n-w, k-1, d']$ code, with $d' \geq d - w + \left\lceil \frac{w}{q} \right\rceil$

Cor. U7#3.3. If code is $[n, k, d]$ code and punctured on codeword weight d, residual code is $[n-d, k-1, d']$ code with $d' \geq \left\lceil \frac{d}{q} \right\rceil$

Thm U7#3.4. (**Griesmer**)

$$n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil$$

Thm U7#3.7. Every non-zero codeword of the r-dimensional simplex code over $GF(q)$ has weight q^{r-1} ("**constant weight code**")

Ex U7#3.8. These codes satisfy the Griesmer bound with equality

---

Thm U8#1.3. For a symmetric encryption scheme $\boxed{\text{with } |K| = |C| = |M|}$: perfect secrecy $\iff$
- each key is chosen with probability $\frac{1}{|K|}$
- every $m \in M, c \in C$, there is unique key K with $\text{Enc}(k, m) = c$

Thm U8#2.2. Let π be the period of the sequence output by an r-bit LFSR: $\pi \leq 2^r - 1$

Thm U8#2.4. LFSR with taps corresponding to degree r primitive polynomial f with $f(\theta) = 0$: The set of sequences output by the LFSR form an r-dimensional vector space
$\implies$ and can be treated as a linear code

Page U8#7. Specifically, a **simplex** code (see immediately from fib LFSR)
$\mathbf{m}$ as initial state (r bits), m-sequence as codeword ($2^r - 1$ bits)

Ex U8#2.6. Fib LFSR satisfies recurrence relation // every m-sequence output by a Galois LFSR can also be generated by a Fib LFSR

Thm U8#2.8. Properties of m-sequence output by r-bit LFSR:
- coordinates of each non-zero length r vector occur exactly once as r consecutive terms of the sequence (think of Fib LFSR internal states)
- number of runs of i consecutive 1s bzw. 0s:
2^{r-i-2}
$i = r - 1$: no runs of 1s / 1 run of 0s
$i = r$: 1 run of 1s / no runs of 0s
- Autocorrelation of the m-seq. is either -1 or $2^r - 1$

Thm U9#1.3. (**Shamir's threshold scheme**) To construct a $(k; n)$ threshold scheme;
$q = p^r; q > n; s \in GF(q)$
Pick $f(x)$ polynomial degree $k - 1$ (or less) with coefficients in $GF(q)$ / constant term s

---

Page U9#5. MDS code $\rightarrow$ secret sharing scheme // rows as distribution rules (Theorem(?): MDS code always yields $(k; n)$ threshold scheme)

Thm U9#1.10. Perfect secret sharing scheme: information rate ≤ 1

Ex U9#1.11. $\text{H}(\mathbf{K}) \leq H(\mathbf{S}_i)$

Thm U9#1.19. Given $[n, m, d]$ code with dual dist. d^*, then for any $1 \leq s \leq d^* - 2$, there exists a $(t_1, t_2, n - s)$ ramp scheme with $t_1 = d^* - s - 1$ and $t_2 = n - d + 1$

Ex U9#1.21. Define average information rate $= \dfrac{n \log_2(|K|)}{\log_2(|S|)}$: for any (t_1, t_2, n) ramp scheme this rate is at most $t_2 - t_1$

Thm U10#1.5. Let $\mathcal{C} \subseteq Q^n$ be a q-ary length n code with M codewords. If $M - 1 \geq c \geq q$ then $\mathcal{C}$ is not a c-TA code.

Thm U10#1.6. A q-ary length n c-TA code $\mathcal{C}$ satisfies

$$|\mathcal{C}| \leq q^{\lceil \frac{n}{c} \rceil} + 2c - 2$$

Thm U10#1.7. For $c \geq 2$, a q-ary length n code with min dist. d and $d > n - \lceil \frac{n}{c^2} \rceil$ is a c-TA code.

Page U10#5. A c-TA code is also a c-frameproof code

Thm U10#1.11. A $[n, \left\lceil \dfrac{n}{c} \right\rceil, n - \left\lceil \dfrac{n}{c} \right\rceil + 1]$ RS code is a c-FP code for $c \geq 2$

Thm U10#1.12. A q-ary length n c-FP code $\mathcal{C}$ with $c < n$ satisfies

$$|\mathcal{C}| \leq \max\{q^{\lceil \frac{n}{c} \rceil}, t(q^{\lceil \frac{n}{c} \rceil} - 1) + (c - t)(q^{\lfloor \frac{n}{c} \rfloor} - 1)\}$$

where t is the remainder when n is divided by c

Ex U10#1.13. for any $c \geq n$, the set of elements of $\{0, 1, \ldots, q - 1\}^n$ with exactly one non-zero component is a c-FP code with $n(q-1)$ elements and is (see Thm 1.12) the largest possible c-FP code with these parameters

Probability: d.r.v. $\to$ distribution, joint, conditional, Bayes

entropy: def, properties, conditional, joint etc.;

(fundamental lemma) *from U5: Chebyshev, Fano, ${}^nC_r$ lemma*

mutual information

Channel encoding $\underbrace{\text{instantaneous, u.d.,}}_{\Longrightarrow \text{ compact}}$ $\underbrace{\text{noiseless, noisy}}_{\text{Shannon thms}}$ $\Downarrow$

Huffman $\leftarrow$ *existence, inequalities*

$\to$ (noisy) decoding: ideal obs vs max likelihood

capacity, rate, ...

as channel extension

Block codes Hamming dist, min. dist, error corr., *puncturing, support, residual code*

$\Downarrow$

Linear codes: gen/par mats $\Longrightarrow$ properties

$\Longleftrightarrow$ dual codes —— orthogonal fun

☆ Syndrome decoding ☆

$\rightsquigarrow$ Use of vector space features =

sizes, lin. independence, combo properties etc.

= over a vector space

= over a field:

see GF content

& polys in GFs

Bounds

(egs are linear but bounds are not specific to linear codes)

$A_q \le \dots$

$\le \dots$

$\ge \dots$

$n \ge$

egs = Hamming

Simplex

R-S

as

☆ LFSRs $\to$ as linear codes

$\rightsquigarrow$ *properties*

Security: definitions, Kerckhoff principle, basic ciphers

	Ciphers // encryption	Perfect secrecy
MDS/linear error.corr codes as	Secret-sharing	threshold schemes, ramp schemes
	Piracy	c-TA, c-FP

Thm U1#1.8. (Bayes)

$$\Pr(x|y) = \frac{\Pr(y|x)\Pr(x)}{\Pr(y)}$$

$GF(p^m)/P$ an irreducible polynomial order m is a field of size p^m i.e. has p^m elements

- its cyclic group has $p^m - 1$ elements of which $\phi(p^m - 1)$ are primitive.
- m of them are generated by each primitive polynomial and so there are $\frac{\phi(p^m-1)}{m}$ primitive polynomials (multiple roots not allowed for prim polys)
- to detect if an element is primitive, look at the prime divisors $k_1, k_2, \ldots$ of $p^m - 1$ and try $\frac{p^m-1}{k_i}$ to see if $e^{k_i} = 1$

Ex U1#2.6. no. of prim. elts of $GF(q = p^m)$ is number of things coprime to $q-1$; no. of prim polys is $\frac{\text{prim elts}}{m}$

Ex U1#2.7. Char $p \implies (a+b)^p = a^p + b^p$

Ex U1#2.8. $GF(q^n)$ is a vector space over $GF(q)$ (any prime power q)

Thm U2#2.7. Shannon entropy Any function satisfying Shannon entropy properties (cts, $H_{n+1} \geq H_n$; $0 \leq H \leq \log n$; $\mathrm{H}(\mathbf{X}, \mathbf{Y}) = \mathrm{H}(\mathbf{X}) + \mathrm{H}(\mathbf{X}|\mathbf{Y})$) is a constant multiple of Shannon entropy

---

Thm U3#2.15. (Shannon's Noiseless Coding)

$$\frac{\mathrm{H}(\mathbf{W})}{\log D} \underbrace{\leq}_{\text{Any u.d}} \bar{n} \underbrace{\leq}_{\text{u.d. must exist}} \frac{\mathrm{H}(\mathbf{W})}{\log D} + 1$$

Ex U4#2.11. Hamming dist is a **metric**

Thm U5#1.1. For the binary symmetric channel, NN decoding is equivalent to max. likelihood decoding

---

Thm U5#1.4. NN decoding of a block code with min. dist d can correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors

---

Thm U5#2.2. Capacity of binary symmetric channel with error prob p:

$$1 + p \log p + (1-p) \log (1-p)$$

Page U5#5. If channel has cap. C, its nth extension has cap nC

Ex U9#1.11. $\mathrm{H}(\mathbf{K}) \leq H(\mathbf{S}_i)$

Thm U9#1.19. Given $[n, m, d]$ code with dual dist. d^*, then for any $1 \leq s \leq d^* - 2$, there exists a $(t_1, t_2, n-s)$ ramp scheme with $t_1 = d^* - s - 1$ and $t_2 = n - d + 1$

Ex U9#1.21. Define average information rate $= \dfrac{n \log_2(|K|)}{\log_2(|S|)}$: for any (t_1, t_2, n) ramp scheme this rate is at most $t_2 - t_1$

Ex U10#1.13. for any $c \geq n$, the set of elements of $\{0, 1, \ldots, q-1\}^n$ with exactly one non-zero component is a c-FP code with $n(q-1)$ elements and is (see Thm 1.12) the largest possible c-FP code with these parameters