Information and Communication Theory

Lecture 5

Channels Capacity and Channel Coding

Mário A. T. Figueiredo

DEEC, Instituto Superior Técnico, University of Lisbon, **Portugal**

2023

Channel Coding

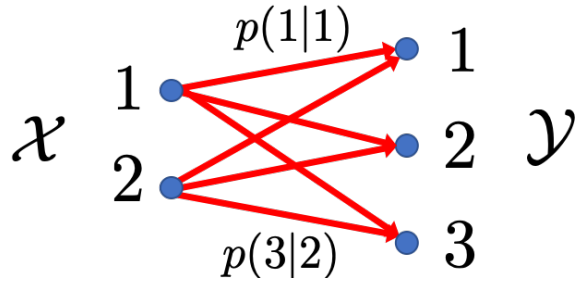
- Message $W \in \{1, ..., M\}$
- Channel input alphabet $\mathcal{X}$; output alphabet $\mathcal{Y}$.
- Encoder: $f : \{1, ..., M\} \to \mathcal{X}^n$.
- Decoder: $g : \mathcal{Y}^n \to \{1, ..., M\}$.
- Message estimate: $\widehat{W} = g(Y^n)$
- Memoryless channel model: $(\mathcal{X}, p(y|x), \mathcal{Y})$

$$p(\underbrace{y_1, ..., y_n}_{y^n} | \underbrace{x_1, ..., x_n}_{x^n}) = \prod_{i=1}^{n} p(y_i|x_i)$$

- An (M, n) code: $(\{1, ..., M\}, f, g)$

Channels

- Memoryless channel model: $(\mathcal{X}, p(y|x), \mathcal{Y})$

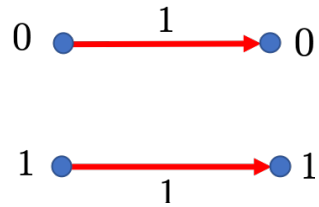


- Channel matrix: $|\mathcal{X}| \times |\mathcal{Y}|$, with $\mathbf{P} = [P_{i,j}] = p(Y = j|X = i)$.
- Channel capacity:

$$C = \max_{p(x)} I(X;Y)$$

the maximum mutual information over all input distributions.

- Example: noiseless binary channel $(Y = X)$:



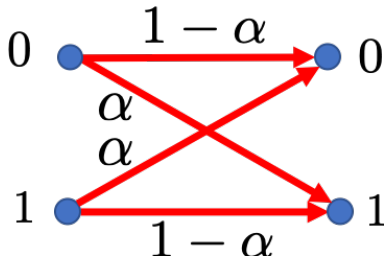
$I(X;Y) = H(Y) = H(X)$.

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) = 1 \text{ bit/symbol}$$

- A noiseless binary channel, can transmit 1 bit/symbol.

Binary Symmetric Channel

- Binary symmetric channel: $(\mathcal{X}, p(y|x), \mathcal{Y})$

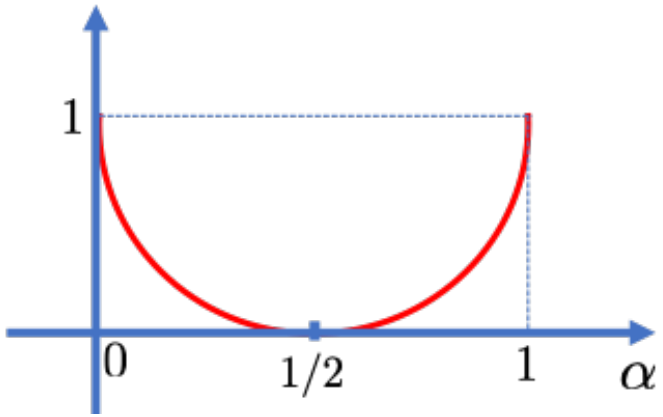


- $H(Y|X = x) = H(\alpha, 1-\alpha)$, for $x = 0$ or $x = 1$.
- $H(Y|X) = H(\alpha, 1-\alpha)$
- $I(X;Y) = H(Y) - H(\alpha, 1-\alpha)$
- Capacity: let $\mathbb{P}(X = 0) = \beta$

$$C = \max_{\beta} H(Y) - H(\alpha, 1-\alpha)$$

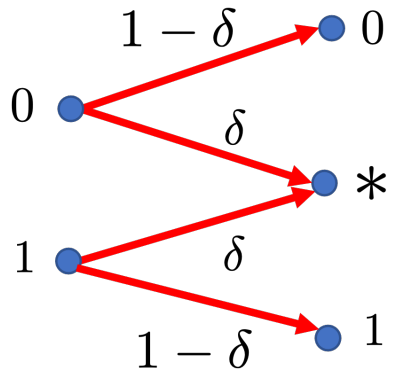
$$= 1 - H(\alpha, 1-\alpha) \text{ bits/symbol}$$

...achieved for $\beta = 1/2$.



- For $\alpha = 0$ or $\alpha = 1$, $C = 1$; for $\alpha = 1/2$, $C = 0$.

Binary Erasure Channel

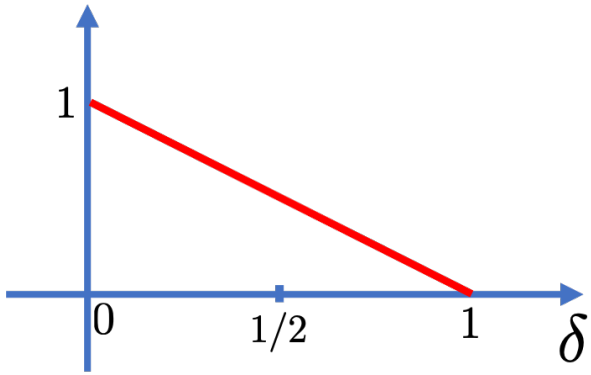


- Binary erasure channel: $(\mathcal{X}, p(y|x), \mathcal{Y})$
- $H(X|Y=1)=0$, for $y=0$ or $y=1$. $H(X|Y=*)=1$
- $H(X|Y)=H(X|Y=*)\mathbb{P}[Y=*]=\delta$
- $I(X;Y)=H(X)-H(X|Y)=H(X)-\delta$
- Capacity: let $\mathbb{P}(X=0)=\beta$

$$C=\max_{\beta} H(X)-\delta$$

$$=1-\delta \text{ bits/symbol}$$

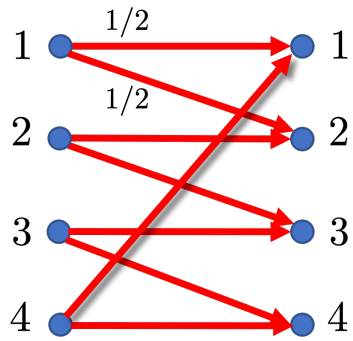
...achieved for $\beta=1/2$.



- For $\delta=0$, $C=1$; for $\delta=1$, $C=0$.

Noisy Typewriter

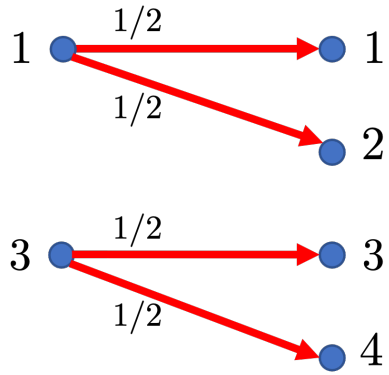
- Noisy typewriter: $(\mathcal{X}, p(y|x), \mathcal{Y})$



- $H(Y|X=x) = 1$, for $x = 1, 2, 3, 4$. $H(Y|X) = 1$
- $I(X;Y) = H(Y) - H(Y|X) = H(Y) - 1$
- Capacity:

$$C = \max_{p(x)} H(Y) - 1$$
$$= 1 \text{ bit/symbol}$$

...achieved for $p(x)$ uniform.



Properties of Channel Capacity

- Because $I(X;Y) \geq 0$, then $C \geq 0$
- Because $I(X;Y) = H(X) - H(X|Y) \leq H(X)$,

$$C = \max_{p(x)} I(X,Y) \leq \max_{p(x)} H(X) \leq \log |\mathcal{X}|$$

- Because $I(X;Y) = H(Y) - H(Y|X) \leq H(Y)$,

$$C = \max_{p(x)} I(X,Y) \leq \max_{p(x)} H(Y) \leq \log |\mathcal{Y}|$$

- Corollary: $C \leq \min\{\log |\mathcal{X}|, \log |\mathcal{Y}|\}$

Channel Coding

- Conditional probability of error (for $i \in \{1, ..., M\}$)

$$\begin{aligned}\lambda_i &= \mathbb{P}\big(g(Y^n) \neq i | X^n = f(i)\big) \\ &= \sum_{y^n \in \mathcal{Y}^n} \mathbb{P}(Y^n = y^n | X^n = f(i))\, \mathbf{1}_{g(y^n) \neq i}\end{aligned}$$

  where $\mathbf{1}_A = 1$, if A is true, and $\mathbf{1}_A = 0$, if A is false.

- Maximum probability of error: $\lambda^{(n)} = \max_{i \in \{1,...,M\}} \lambda_i$
- Average probability of error: $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$
- Probability of error: $\mathbb{P}(g(Y^n) \neq i)$
- If the message symbols are equiprobable: $\mathbb{P}(g(Y^n) \neq i) = P_e^{(n)}$
- Of course, $P_e^{(n)} \leq \lambda^{(n)}$ and $\mathbb{P}(g(Y^n) \neq i) \leq \lambda^{(n)}$

Channel Coding

- The rate of an (M, n) code

$$R = \frac{\log_2 M}{n}$$

- A rate R is achievable if there is a sequence of $(\lceil 2^{nR} \rceil, n)$ codes, s.t.

$$\lim_{n \to \infty} \lambda^{(n)} = 0$$

- (Operational) capacity of a channel is

$$C^{\text{oper}} = \sup\{R : R \text{ is achievable}\}$$

- The channel coding theorem essentially states that

$$C^{\text{oper}} = C = \max_{p(x)} I(X;Y)$$

Channel Coding: Examples

- Consider a quarternary source $M = 4$, thus $\log_2 M = 2$ bits
- Using a binary noiseless channel, $C = 1$ bits/transmission, we need $n = 2$ transmissions to send each symbol:

$$R = \frac{\log_2 M}{n} = \frac{2}{2} = 1 \quad \text{bits/transmission}$$

  is this rate achievable? Yes, because the channel is noiseless.
- What if $C < 1$? Rate 1 is no longer achievable!
- Using a noiseless quarternary channel ($|\mathcal{X}| = |\mathcal{Y}| = 4$): only need $n = 1$ transmissions,

$$R = \frac{\log_2 M}{1} = 2 \quad \text{bits/transmission}$$

  is this rate achievable? Yes, because the channel is noiseless: $C = 2$

Channel Coding: Example

- Consider a binary symmetric channel with $\alpha = 1/5$, thus $C = 0.21$
- Thus, $R = 0.25$ is not achievable; $R = 0.2$ is achievable.
- Examples of sequences $(\lceil 2^{nR} \rceil, n)$, for $R = 0.2$:

  $(2^1, 5), (2^2, 10), (2^3, 15), \ldots$ *e.g.*, use 10-bit codewords to send 2 bits

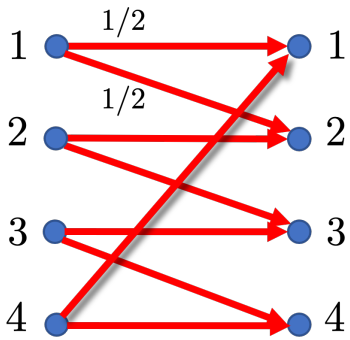
- Examples of sequences $(\lceil 2^{nR} \rceil, n)$, for $R = 0.25$:

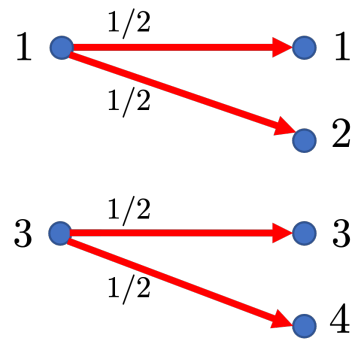
  $(2^1, 4), (2^2, 8), (2^3, 12), \ldots$ *e.g.*, use 8-bit codewords to send 2 bits

- The channel coding theorem states that
  - ✓ There is a sequence of $(\lceil 2^{n\,0.2} \rceil, n)$ codes, s.t. $\lim_{n\to\infty} \lambda^{(n)} = 0$.
  - ✓ For any sequence of $(\lceil 2^{n\,0.25} \rceil, n)$ codes, $\lim_{n\to\infty} \lambda^{(n)} \neq 0$.

Channel Coding: Example

- The noisy typewriter is a simple example of the theorem.





- Capacity: $C = 1$ bit/transmission.
- Input and output alphabets $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4\}$.
- In this case, $R = C = 1$ is achievable.
- Codes $(\lceil 2^n \rceil, n)$ have $\lambda^{(n)} = 0$, for any n.
- Encoder (for $n = 1$, thus $M = 2^n = 2$): $f(1) = 1$; $f(2) = 3$.
- Decoder: $g(1) = g(2) = 1$; $g(3) = g(4) = 2$.
- Since $\lambda^{(n)} = 0$, $C = 1$ is the zero-error capacity.

Asymptotic Equipartition: Motivation 1

- Toss a fair coin 100 times: $X_1, ..., X_{100} \in \{0, 1\}$.
- ...there are $2^{100} \simeq 1.27 \times 10^{30}$ possible outcomes,
- ...each with probability $2^{-100} \simeq 7.9 \times 10^{-30}$
- The overwhelming majority has close to $50/50$ heads/tails ratio.
- How overwhelming? Let $S = X_1 + \cdots + X_{100}$,

$$\mathbb{P}(S \in \{47, ..., 53\}) = 2^{-100} \sum_{j=47}^{53} \binom{100}{j} \simeq 0.52$$

- How many sequences are in this set?

$$|\{(x_1, ..., x_{100}) : S \in \{47, ..., 53\}\}| = \sum_{j=47}^{53} \binom{100}{j} \simeq 6.54 \times 10^{29}$$

...fraction of the total: $\simeq 6.54 \times 10^{29}/2^{100} \simeq 0.52$

Asymptotic Equipartition: Motivation 2

- Unfair coin ($\mathbb{P}(\text{heads}) = \mathbb{P}(X_i = 1) = 0.9$) 100 tosses: X_1, ..., X_{100}.
- ...there are $2^{100} \simeq 10^{30}$ possible outcomes,
- The overwhelming majority has close to $90/10$ heads/tails ratio.
- How overwhelming? Let $S = X_1 + \cdots + X_{100}$,

$$\mathbb{P}(S \in \{87, ..., 93\}) = \sum_{j=87}^{93} 0.9^j \, 0.1^{100-j} \binom{100}{j} \simeq 0.76$$

- How many sequences are in this set?

$$|\{(x_1, ..., x_{100}) : S \in \{87, ..., 93\}\}| = \sum_{j=87}^{93} \binom{100}{j} \simeq 8.3 \times 10^{15}$$

...fraction of the total: $\simeq 8.3 \times 10^{15}/2^{100} \simeq 6.5 \times 10^{-15}$

Asymptotic Equipartition: Law of Large Numbers

- Consider X_1,...,X_n i.i.d. with $\mathbb{E}[X_i] = \mu$.
- Weak law of large numbers (WLLN) (Bernoulli, 1713)

$$\lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) = \mu, \quad \text{(in probability)}$$

- Applying to $\log p(X_1, ..., X_n)$,

$$-\frac{1}{n}\log p(X_1, ..., X_n) = -\frac{1}{n}\sum_{i=1}^{n} \log p(X_i) \underset{n\to\infty}{\longrightarrow} \mathbb{E}[-\log p(X)] = H(X)$$

- This convergence is in probability: for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left|-\frac{\log p(X_1, ..., X_n)}{n} - H(X)\right| < \varepsilon\right] = 1$$

Asymptotic Equipartition Property (AEP)

- Definition: for n i.i.d. samples $x_1, ..., x_n$ of $X \in \mathcal{X}$, the set of ε-typical sequences $A_\varepsilon^{(n)}$ (called typical set) is

$$A_\varepsilon^{(n)} = \{(x_1, ..., x_n) : \left| -\frac{1}{n} \log p(x_1, ..., x_n) - H(X) \right| \leq \varepsilon\}$$

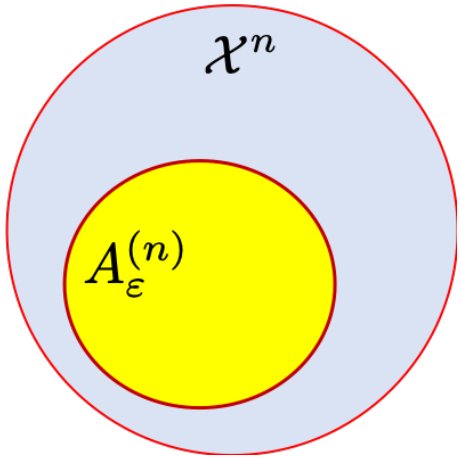
- The condition can also be written as

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, ..., x_n) \leq 2^{-n(H(X)-\varepsilon)}$$

- AEP theorem: for any $\varepsilon > 0$ and n sufficiently large,

$$\mathbb{P}\left[(x_1, ..., x_n) \in A_\varepsilon^{(n)}\right] \geq 1 - \varepsilon$$

$$(1-\varepsilon)2^{n(H(X)-\varepsilon)} \leq |A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$



AEP Corollary

- AEP theorem: for any $\varepsilon > 0$ and n sufficiently large,

$$\mathbb{P}\left[(x_1, ..., x_n) \in A_\varepsilon^{(n)}\right] \geq 1 - \varepsilon$$

$$(1-\varepsilon)2^{n(H(X)-\varepsilon)} \leq |A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$

- Corollary: if $H(X) < \log|\mathcal{X}|$, for a sufficiently small ε,

$$\lim_{n\to\infty} \frac{|A_\varepsilon^{(n)}|}{|\mathcal{X}^n|} \leq \lim_{n\to\infty} 2^{n(H(X)+\varepsilon-\log|\mathcal{X}|)} = 0,$$

  if $\varepsilon < \log|\mathcal{X}| - H(X)$.

- Typical set: vanishingly small volume with arbitrarily high probability.

AEP: Example 1

- Toss a fair coin n times: $X_1, ..., X_n \in \{0, 1\}$; $H(X) = 1$
- ...there are 2^n possible outcomes,
- ...each with probability $p(x_1, ..., x_n) = 2^{-n}$
- With $\varepsilon = 0.02$,

$$A_{0.02}^{(n)} = \{(x_1, ..., x_n) : 2^{-n(1.02)} \leq p(x_1, ..., x_n) \leq 2^{-n(0.98)}\} = \{0, 1\}^n$$

$$0.98\, 2^{n(0.98)} \leq |A_{0.02}^{(n)}| \leq 2^{n(1.02)} \qquad (|A_{0.02}^{(n)}| = 2^n)$$

$$1 = \mathbb{P}\left[(x_1, ..., x_n) \in A_{0.02}^{(n)}\right] \geq 0.98$$

- For maximum entropy, $H(X) = 1$, the AEP is uninformative/vacuous.

AEP: Example 2

- Toss a unfair coin (probability of heads 0.8) n times: X_1, ..., X_n.
- Entropy: $H(X) \simeq 0.72$
- With $\varepsilon = 0.02$,

$$A_{0.02}^{(n)} = \{(x_1, ..., x_n) : 2^{-n(0.74)} \leq p(x_1, ..., x_n) \leq 2^{-n(0.70)}\}$$

$$|A_{0.02}^{(n)}| \leq 2^{n\,0.74}$$

$$\mathbb{P}\left[(x_1, ..., x_n) \in A_{0.02}^{(n)}\right] \geq 0.98, \quad \text{for } n \text{ large enough}$$

- For non-maximum entropy, the AEP is very informative:

$$\frac{|A_{0.02}^{(n)}|}{|\{0,1\}^n|} \leq 2^{-n\,0.26} \quad \text{(e.g., for } n = 100, 2^{-26} \simeq 10^{-8}\ )$$

Interlude: AEP and Source Coding

- Source $X \in \mathcal{X}$; order n extension of : $X^n = (X_1, ..., X_n) \in \mathcal{X}^n$.
- Coding method: given a sequence $(x_1, ..., x_n)$,
  - ✓ if $(x_1, ..., x_n) \in A_\varepsilon^{(n)}$ code it using $\lceil n(H(X) + \varepsilon) \rceil$ bits.

    ...enough, because $\#A_\varepsilon^{(n)} \leq 2^{n(H(X)+\varepsilon)}$
  - ✓ if $(x_1, ..., x_n) \notin A_\varepsilon^{(n)}$, code it using $\lceil \log |\mathcal{X}^n| \rceil = \lceil n \log |\mathcal{X}| \rceil$ bits.

    ...enough, because $|\bar{A}_\varepsilon^{(n)}| \leq |\mathcal{X}^n| = |\mathcal{X}|^n$
  - ✓ to distinguish the two cases, use a 1-bit prefix.
- Length of this coding scheme:

$$l_C(x_1, .., x_n) = \begin{cases} 1 + \lceil n(H(X) + \varepsilon) \rceil & \text{if } (x_1, ..., x_n) \in A_\varepsilon^{(n)} \\ 1 + \lceil n \log |\mathcal{X}| \rceil & \text{if } (x_1, ..., x_n) \notin A_\varepsilon^{(n)} \end{cases}$$

Interlude: AEP and Source Coding

- Length of the coding scheme in the previous slide:

$$l_C(x_1,..,x_n) < \begin{cases} 2 + n(H(X)+\varepsilon) & \text{if } (x_1,....,x_n) \in A_\varepsilon^{(n)} \\ 2 + n\log|\mathcal{X}| & \text{if } (x_1,....,x_n) \notin A_\varepsilon^{(n)} \end{cases}$$

- Expected length $L[C] = \mathbb{E}[l_C(X_1,...,X_n)]$ (in bits/(n symbols)), for $0 < \varepsilon \ll 1$ and sufficiently large n,

$$\begin{aligned} L[C] &< \underbrace{\mathbb{P}[A_\varepsilon^{(n)}]}_{\leq 1}\,(2 + n(H(X)+\varepsilon)) + \underbrace{\big(1 - \mathbb{P}[A_\varepsilon^{(n)}]\big)}_{\leq \varepsilon}\,(2 + n\log|\mathcal{X}|) \\ &\leq 2 + (n(H(X)+\varepsilon)) + \varepsilon(n\log|\mathcal{X}|) \\ &= 2 + n\big(H(X) + \varepsilon + \varepsilon\log|\mathcal{X}|\big) \end{aligned}$$

- Normalize to bits/symbol, dividing by n:

$$\frac{L[C]}{n} \leq \frac{2 + n\big(H(X) + \varepsilon + \varepsilon\log|\mathcal{X}|\big)}{n},$$

...that is, $L[C]/n$ can be arbitrarily close to $H(X)$.

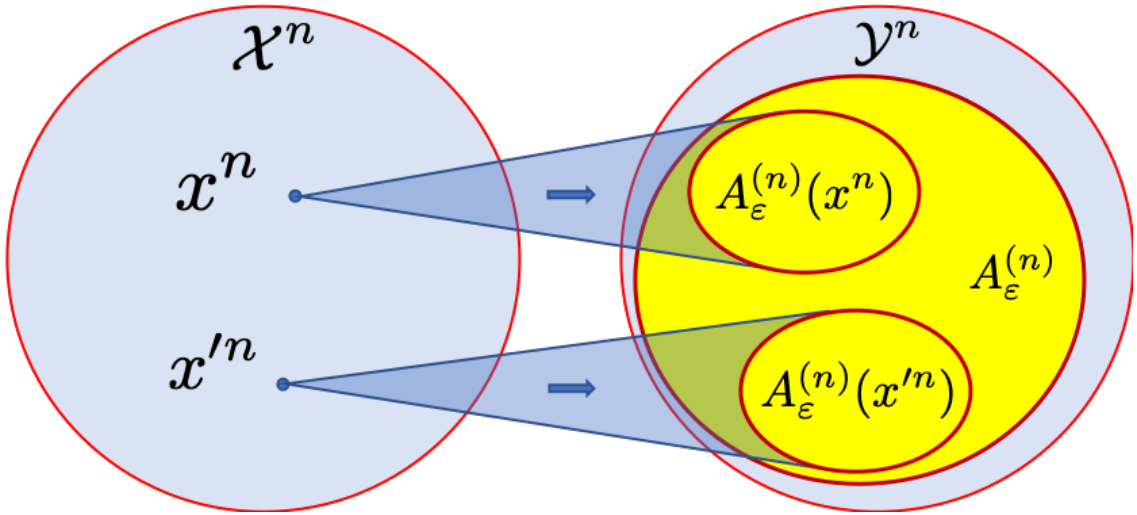
Channel Coding Theorem

- Consider a discrete memoryless channel with capacity C. Then,

  1) Any $R < C$ is achievable: there exist sequences of $(\lceil 2^{nR} \rceil, n)$ codes such that $\lim_{n\to\infty} \lambda^{(n)} = 0$.

  2) Any sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $\lim_{n\to\infty} \lambda^{(n)} = 0$, must have $R \leq C$

- Intuition: for large n, every channel looks like a noisy typewriter.



Channel Coding Theorem: Overview of the Proof

- Take $p^*(x) = \arg\max_{p(x)} I(X;Y)$, and $p(x^n) = \prod_{i=1}^n p^*(x_i)$
- For every x^n, consider $H(Y^n|X^n = x^n)$; conditional typical set

$$A_\varepsilon^{(n)}(x^n) = \{y^n : \left|-\frac{1}{n}\log p(y^n|x^n) - H(Y|X)\right| \leq \varepsilon\}$$

- For arbitrarily small ε and large n, AEP states that

$$|A_\varepsilon^{(n)}(x^n)| \simeq 2^{nH(Y|X)}$$

- The unconditional typical set is

$$A_\varepsilon^{(n)} = \{y^n : \left|-\frac{1}{n}\log p(y^n) - H(Y)\right| \leq \varepsilon\}$$

- For arbitrarily small ε and large n, AEP states that

$$|A_\varepsilon^{(n)}| \simeq 2^{nH(Y)}$$

Channel Coding Theorem: Overview of the Proof

- To have (asymptotically as $n \to \infty$) error-free communication:
  - ✓ Different $A_\varepsilon^{(n)}(x^n)$ must be disjoint:
  - ✓ All $A_\varepsilon^{(n)}(x^n)$ must be inside $A_\varepsilon^{(n)}$
- The maximum number of words that we can have is thus,

$$M = 2^{n\,R} \leq \frac{|A_\varepsilon^{(n)}|}{|A_\varepsilon^{(n)}(x^n)|} \simeq 2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)} \leq 2^{n\,C}$$

  ...which leads to $R < C$.

- This was not a rigorous proof; if you're interested in the details, see the recommended reading.

Repetition Codes

- Unlike for source coding (where we have Huffman codes), building capacity-approaching codes is harder.
- Simplest code: repetition; e.g., $(\lceil 2^{n/3} \rceil, n)$ codes, rate $R = 1/3$.
- For $n = 3$, we have $(2, 3)$-codes, thus $M = 2$ words, $W \in \{0, 1\}$,

$$\text{encoder } f(0) = 000, f(1) = 111$$
$$\text{decoder } g(y^3) = \arg \min_{i \in \{0,1\}} d_H(y^3, f(i)),$$

  where d_H is the Hamming distance (number of bits in which the words differ): minimum distance decoding.
- For higher n, we have $(2^2, 6)$ codes $(M = 4)$, ..., $(2^5, 15)$ codes $(M = 32)$, ...

Error Correction and Error Detection

- A binary encoder $f : \{1, ..., M\} \to \{0, 1\}^n$ defines a set of codewords:

$$\{f(1), f(2), ..., f(M)\}.$$

- Minimum distance decoding of received word y^n:

$$g(y^n) = \arg \min_{i \in \{1,...,M\}} d_H(y^n, f(i))$$

- Minimum distance of the code:

$$d_{\text{min}} = \min_{i \neq j} d_H(f(i), f(j))$$

- Error correction: a code corrects up to $\frac{d_{\text{min}} - 1}{2}$ errors.
- Error detection: a code detects up to $d_{\text{min}} - 1$ errors.
- Exercise: show that a repetition code corrects up to $\frac{1-R}{2R}$ errors and detects up to $(1 - R)/R$ errors.

Hamming Codes

- Binary linear codes are built on binary linear algebra.
- Before proceeding, we need need binary arithmetic: $(\{0,1\},+,\times)$
  - ✓ addition: $0+0=0$, $0+1=1$, $1+0=1$, $1+1=0$.
  - ✓ multiplication: $0\times 0=0$, $0\times 1=0$, $1\times 0=0$, $1\times 1=1$.
  - ✓ both are clearly commutative $a+b=b+a$ and $a\times b=b\times a$.
  - ✓ also associative: $a+(b+c)=b+(a+c)$ and $(a\times b)\times c=a\times(a\times c)$.
  - ✓ distributive property: $a\times(b+c)=a\times b+a\times c$.
- In binary arithmetic, $a+b=a-b$.
- Based on binary arithmetic, we may build binary linear algebra, with binary vectors and matrices.
- Can be extended to other Galois fields $GF(q)$; *e.g.*, $GF(q)$ ternary arithmetic, with modulo-3 addition and multiplication.

Hamming Codes

- Generalizes the idea of parity check for error detection/correction.
- Hamming(n, k) code is (in the previous notation) a $(2^k, n)$ code.
- Rate of a Hamming(n, k) code: $R = k/n$.
- Classical example: Hamming$(7, 4)$ generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} = \left[\mathbf{I}_4 | \mathbf{A}\right]$$

- Generation of codeword $\mathbf{x}$ from message: example $\mathbf{m} = (1101)$:

$$\mathbf{x} = \mathbf{mG} = (1101)\mathbf{G} = (1101100)$$

where the vector-matrix product is in binary arithmetic.

Hamming Codes

- Generation of codeword $\mathbf{x}$ from message: example $\mathbf{m} = (1101)$:

$$\mathbf{x} = \mathbf{mG} = (1101)\mathbf{G} = (1101100)$$

- Checking codewords:parity-check matrix $\mathbf{H}$ such that

$$\mathbf{HG}^T = \mathbf{0} \;\Rightarrow\; \mathbf{Hx}^T = \mathbf{H}(\mathbf{mG})^T = \mathbf{HG}^T\mathbf{m}^T = \mathbf{0}$$

- For $\mathbf{G} = \left[\mathbf{I}_4|\mathbf{A}\right]$, then $\mathbf{H} = \left[\mathbf{A}^T|\mathbf{I}_3\right]$

$$\mathbf{HG}^T = \left[\mathbf{A}^T|\mathbf{I}_3\right]\begin{bmatrix}\mathbf{I}_4\\ \mathbf{A}^T\end{bmatrix} = \mathbf{A}^T + \mathbf{A}^T = 0$$

- For matrix $\mathbf{G}$ in the previous slide

$$\mathbf{H} = \left[\mathbf{A}^T|\mathbf{I}_3\right] = \begin{bmatrix}1 & 1 & 0 & 1 & 1 & 0 & 0\\ 1 & 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 1 & 1 & 1 & 0 & 0 & 1\end{bmatrix}$$

...the columns are the $(2^3 - 1 = 7)$ 3-bit binary words, except (000).

Hamming Codes

- Let $\mathbf{x} + \mathbf{e}$ be a received codeword, with error vector $\mathbf{e}$.
- Checking: $\mathbf{H}(\mathbf{x} + \mathbf{e})^T = \underbrace{\mathbf{H}\mathbf{x}^T}_{0} + \mathbf{H}\mathbf{e}^T = \mathbf{H}\mathbf{e}^T$
- No errors detected if and only if $\mathbf{H}\mathbf{e}^T = 0$. Conditions:
  - ✓ Zero errors, $\mathbf{H}\mathbf{e}^T = 0$.
  - ✓ One error, $\mathbf{H}\mathbf{e}^T \neq 0$; it is one of the columns of $\mathbf{H}$.
  - ✓ Two errors, $\mathbf{H}\mathbf{e}^T \neq 0$; it is the sum of two columns of $\mathbf{H}$, which are all different.
- Any two errors are detected, but three errors may be undetected, since the sum of any two columns equals another column.

Hamming Codes

- Minimum distance of Hamming$(7, 4)$ code is 3.
- Thus it can correct 1 error; how?
- Permute the columns of $\mathbf{H}$ (and similarly of $\mathbf{G}$) into

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- Check $\mathbf{x} + \mathbf{e}$, assuming only one error in position, say 5,

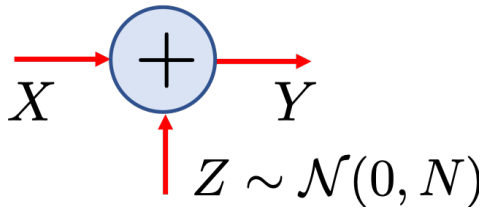
$$(\mathbf{H}(\mathbf{x}^T + \mathbf{e}^T))^T = \mathbf{e}\mathbf{H}^T = (0000100)\mathbf{H}^T = (101)$$

...precisely the binary word for 5, the error position.

Hamming Codes

- General Hamming(n, k) codes.
- For some $r \geq 2$: $n = 2^r - 1$ and $k = 2^r - r - 1$.
- The Hamming$(7, 4)$ code: $r = 3$, $n = 2^3 - 1 = 7$, $k = 2^3 - 3 - 1 = 4$.
- Columns $\mathbf{H}$: all $n = 2^r - 1$ binary words of r bits, except zero.
- Put $\mathbf{H}$ is systematic form $\mathbf{H} = \left[\mathbf{A}^T | \mathbf{I}_3\right]$ and build $\mathbf{G} = \left[\mathbf{I}_4 | \mathbf{A}\right]$.
- Rate: $R = k/n = (2^r - r - 1)/(2^r - 1)$
- Exercise: show that, for any r, $d_{\text{min}} = 3$.
- Remarkably, $\lim_{r \to \infty} \dfrac{2^r - r - 1}{2^r - 1} = 1$
- The repetition code also has minimum distance 3, but $R = 1/3$.
- Error-correcting codes are a huge R&D area, without which modern communications would not be possible.

Gaussian Channel



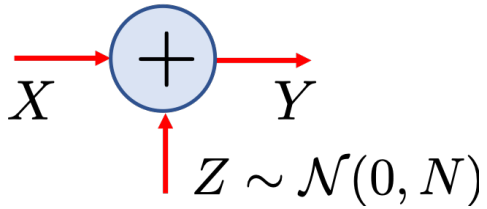
- Gaussian channel: $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $Y = X + Z$.
- Mutual information

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z) \\ &= h(Y) - \frac{1}{2}\log\left(2\pi e N\right) \end{aligned}$$

since differential entropy is shift-invariant, and Z is Gaussian and independent of X.

- Since adding a constant to X does not affect $I(X;Y)$, assume $\mathbb{E}[X] = 0$, thus $\text{var}[X] = \mathbb{E}[X^2] =$ power.

Gaussian Channel



- Gaussian channel: $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $Y = X + Z$.
- Mutual information,

$$\begin{aligned} I(X;Y) &= h(Y) - \frac{1}{2}\log\big(2\pi e N\big) \\ &\leq \frac{1}{2}\log\big(2\pi e(N + \mathbb{E}[X^2])\big) - \frac{1}{2}\log\big(2\pi e N\big) \\ &= \frac{1}{2}\log\Big(1 + \frac{\mathbb{E}[X^2]}{N}\Big) \end{aligned}$$

- Without a constraint on $\mathbb{E}[X^2]$, $I(X;Y)$ is arbitrarily large.
- With a power constrain $\mathbb{E}[X^2] \leq P$,

$$C = \max_{f_X : \mathbb{E}[X^2] \leq P} \frac{1}{2}\log\Big(1 + \frac{\mathbb{E}[X^2]}{N}\Big) = \frac{1}{2}\log\Big(1 + \frac{P}{N}\Big)$$

achieved for $f_X = \mathcal{N}(0, P)$. $P/N = SNR$, signal to noise ratio.

Coding for a Gaussian Channel

- An (M, n) code for a Gaussian channel, under power constraint P.
  - ✓ a set of message indices $W \in \{1, ..., M\}$;
  - ✓ an encoder $f : \{1, ..., M\} \to \mathbb{R}^n$, *i.e.*, $f(i) = [f_1(i), ..., f_n(i)] \in \mathbb{R}^n$,

$$\|f(i)\|^2 = \sum_{j=1}^{n} f_j(i)^2 \leq nP.$$

  - ✓ a decoder $g : \mathbb{R}^n \to \{1, ..., M\}$.
- Conditional, average, and maximum probability of error, $\lambda^{(n)}$, are defined as in the discrete channel.
- Rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes satisfying the power constraint P, such that $\lim_{n\to\infty} \lambda^{(n)} = 0$.
- The (operational) capacity is: $C^{\text{oper}} = \sup\{R : R \text{ is achievable}\}$.
- The Gaussian channel theorem: $C^{\text{oper}} = C$.

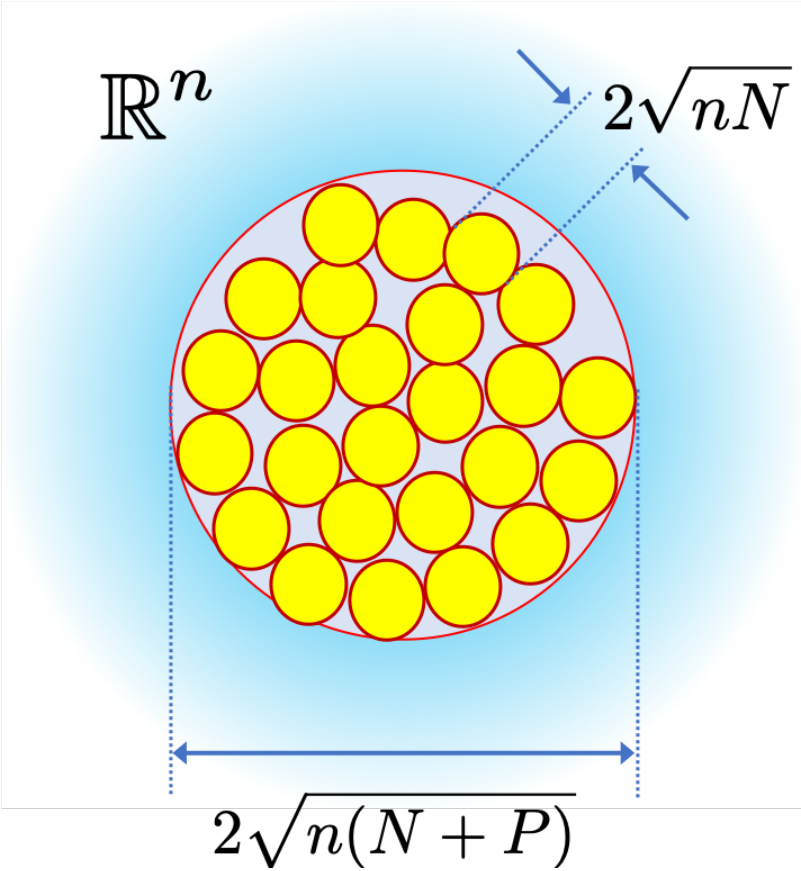
Coding for a Gaussian Channel

- Outline of the proof of the Gaussian channel theorem.
- We known that $Y^n = X^n + Z^n$, that $\mathbb{E}[\|X^n\|^2] \leq nP$, thus $\mathbb{E}[\|Y^n\|^2] \simeq n(P+N)$.
- All the received vectors are, with high probability (w.h.p.), in a sphere of radius $\sqrt{n(N+P)}$.
- Each received vector is, w.h.p., in a sphere around $f(i)$ of radius $\sqrt{nN}$.
- The volume of a radius-r sphere is $V(r) = C_n r^n$.
- The maximum number of (asymptotically) non-intersecting spheres is

$$M = 2^{nR} \leq \frac{(n(N+P))^{n/2}}{(nN)^{n/2}} = 2^{\frac{n}{2}\log\left(\frac{P+N}{N}\right)} = 2^{\frac{n}{2}\log\left(1+\frac{P}{N}\right)}$$

...thus $R < C$.

Sphere Packing



- Thus picture becomes accurate for large n, since "in high dimensions, Gaussian distributions are soap bubbles."[1]

[1] www.inference.vc/high-dimensional-gaussian-distributions-are-soap-bubble/

Recommended Reading

- T. Cover and J. Thomas, "Elements of Information Theory", John Wiley & Sons, 2006 (Sections 7.1 to 7.6, 7.11, 9.1).
- `https://en.wikipedia.org/wiki/Hamming_code`