COMP2610/6261 - Information Theory

Lecture 20: Joint-Typicality and the Noisy-Channel Coding Theorem

Australian National University

October 24th, 2018

Channel Capacity: Recap

The *largest possible* reduction in uncertainty achievable across a channel is its **capacity**

Channe

The capa een its input and output for any choice of input ensemble. That is,

$$C = \max_{\mathbf{p}_X} I(X$$

Block Codes: Recap

(N, K) Block Code

Given a channel Q with inputs and outputs , an integer $N > 0$, and $K > 0$, an ds

where each $\mathbf{x}^{(s)} \in \mathcal{X}^N$ consists of N symb

Rate of a block code is $\frac{K}{N} = \frac{\log_2 S}{N}$

Reliability: Recap

$$\mathbf{s}_{in} \xrightarrow{\text{Encoder}} \mathbf{x} \xrightarrow{Q} \mathbf{y} \xrightarrow{\text{Decoder}} \mathbf{s}_{out}$$

Probabi

Given a ch

$$p_B = P(\mathbf{s}_{out} \neq \mathbf{s}_{in}) = \sum_{\mathbf{s}_{in}} P(\quad / $$

and its **maximum probability of (block) error**

$$p_{BM} = \max_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in})$$

The Noisy-Channel Coding Theorem: Recap

Informal Statement

Recall that a rate R is achievable if there is a block code with this rate and arbitrarily small error probability

We highli

Noisy-C

If Q is a ch **if and only if** $R \leq C$, that is, the rate is no greater than the channe

The Noisy-Channel Coding Theorem: Recap

Informal Statement

Recall that a rate R is achievable if there is a block code with this rate and arbitrarily small error probability

We highli

Noisy-C

If Q is a ch **if and only if** $R \leq C$, that is, the rate is no greater than the channe

Ideally, we would like to know:

- Can we go above C if we allow some fixed probability of error?
- Is there a maximal rate for a fixed probability of error?

1. Noisy-Channel Coding Theorem

2. Joint Typicality

3. Proof

4. Good Codes vs. Practical Codes

5. Linear Codes

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem (Formal)

1. Any r

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem (Formal)

1. Any r
2. If pro codes exis

$$\frac{\quad}{N} \leq R(p_b) = \frac{\quad}{\quad}$$

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem (Formal)

1. Any r
2. If pro codes exis

$$\overline{N} \leq R(p_b) = \overline{\qquad}$$

3. For any p, we **cannot** achieve a rate gre probability of bit error p.

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem (Formal)

1. Any r
2. If pro codes exis

$$\frac{}{N} \leq R(p_b) = \frac{}{}$$

3. For any p, we **cannot** achieve a rate gre probability of bit error p.

The Noisy-Channel Coding Theorem

Formal Statement

Recall: a rate is achievable if for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$

The Noisy-Channel Coding Theorem (Formal)

1. Any r
2. If pro codes exis

$$\frac{\quad}{N} \leq R(p_b) = \frac{\quad}{\quad}$$

3. For any p, we **cannot** achieve a rate gre probability of bit error p.

Note that as $p_b \to \frac{1}{2}$, $R(p_b) \to +\infty$, while as $p_b \to \{0, 1\}$, $R(p_b) \to C$, so we cannot achieve rate greater than C with probability of bit error arbitrarily small

Implications of NCCT

Suppose we know a channel has capacity 0.6 bits

Implications of NCCT

Suppose we know a channel has capacity 0.6 bits

We **cannot** achieve a rate of 0.8 with arbitrarily small error

Implications of NCCT

Suppose we know a channel has capacity 0.6 bits

We **cannot** achieve a rate of 0.8 with arbitrarily small error

We **can** achieve a rate of 0.8 with probability of bit error 5%, since $\frac{0.6}{1-H_2(0.05}$

Joint Typicality

Recall that a random variable $\mathbf{z}$ from Z^N is typical for an ensemble Z whenever its average symbol information is within β of the entropy $H(Z)$

$$\left| -\frac{1}{N} \log_2 P(\mathbf{z}) - H(Z) \right| < \beta$$

Joint Typicality

Recall that a random variable $\mathbf{z}$ from Z^N is typical for an ensemble Z whenever its average symbol information is within β of the entropy $H(Z)$

$$\frac{1}{\ } \log_2 P(\mathbf{z}) \quad H(Z) \ < \beta$$

Joint Ty

A pair of sequences $\mathbf{x} \in \mathcal{A}_X^N$ and $\mathbf{y} \in \mathcal{A}_Y^N$, e **ntly typical** (to tolerance β) for distribution $P(x$

1. $\mathbf{x}$ is typical of $P(\mathbf{x})$ … above]
2. $\mathbf{y}$ is typical of $P(\mathbf{y})$ … $=$ above]
3. $(\mathbf{x}, \mathbf{y})$ is typical of $P(\mathbf{x}, \mathbf{y})$ [$\mathbf{z} = (\mathbf{x}, \mathbf{y})$ above]

The **jointly typical set** of all such pairs is denoted $J_{N\beta}$.

Joint Typicality

Example

Example ($\mathbf{p}_X = (0.9, 0.1)$ and BSC with $f = 0.2$):

$\mathbf{x}$ 1111111111000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

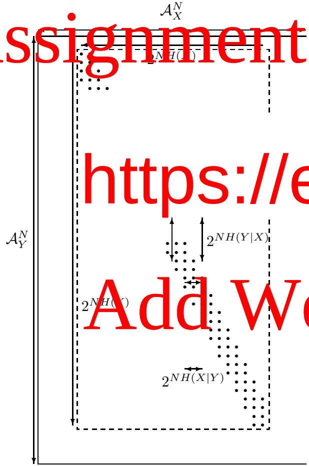
$\mathbf{y}$ 0011111111000000000000000000000000000000000000000000000000000000000000000000000000111111111111111111

Here:

- x has 10 1's (c.f. $p(X = 1) = 0.1$)

- y has 26 1's (c.f. $p(Y = 1) = (0.8)(0.$

- x, y differ in 20 bits (c.f. $p(X \neq Y) = 0.2$)
  - This is essential in addition to the above two facts

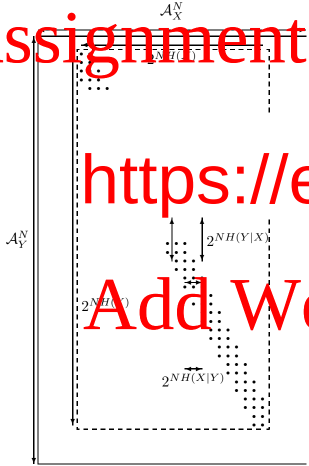


There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$

Joint Typicality

Counts

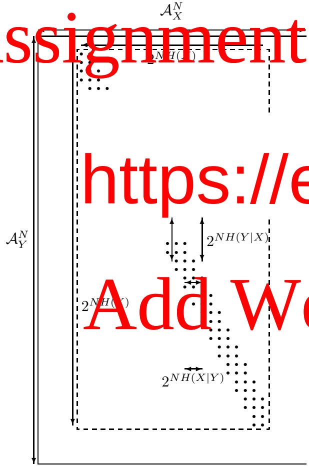


There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$
- $2^{NH(Y)}$ typical $\mathbf{y} \in \mathcal{A}_Y^N$

Joint Typicality

Counts

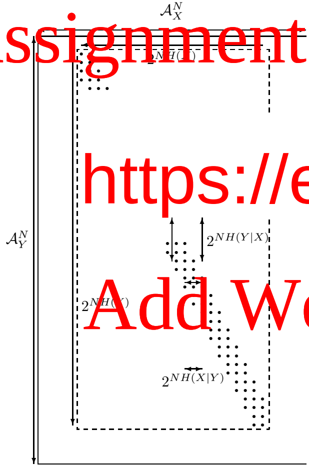


There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$
- $2^{NH(Y)}$ typical $\mathbf{y} \in \mathcal{A}_Y^N$
- $2^{NH(X,Y)}$ typical $(\mathbf{x}, \mathbf{y})$ $\mathcal{A}_X^N \times \mathcal{A}_Y^N$

Joint Typicality

Counts

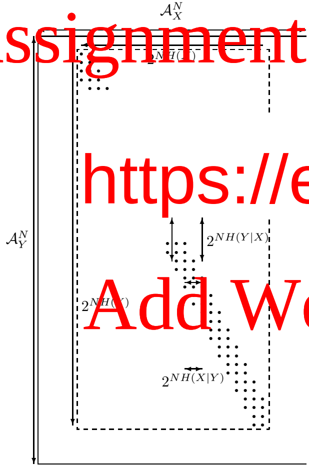


There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$
- $2^{NH(Y)}$ typical $\mathbf{y} \in \mathcal{A}_Y^N$
- $2^{NH(X,Y)}$ typical $(\mathbf{x}, \mathbf{y})$ $_X^N \times \mathcal{A}_Y^N$
-

Joint Typicality

Counts



There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$
- $2^{NH(Y)}$ typical $\mathbf{y} \in \mathcal{A}_Y^N$
- $2^{NH(X,Y)}$ typical $(\mathbf{x}, \mathbf{y})$ $_X^N \times \mathcal{A}_Y^N$
-

ical

vectors, we

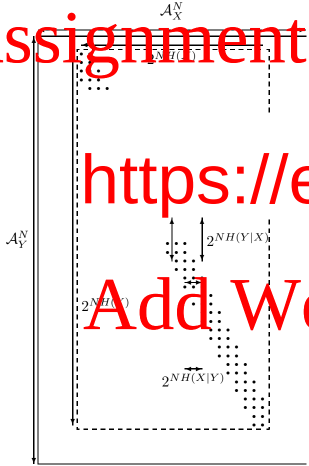
with probab

$$\frac{}{2^{NH(X)} \cdot 2^{NH(Y)}} = 2^{-}$$

(;)

Joint Typicality

Counts



There are approximately:

- $2^{NH(X)}$ typical $\mathbf{x} \in \mathcal{A}_X^N$
- $2^{NH(Y)}$ typical $\mathbf{y} \in \mathcal{A}_Y^N$
- $2^{NH(X,Y)}$ typical $(\mathbf{x}, \mathbf{y})$ $_X^N \times \mathcal{A}_Y^N$
-

ical

vectors, we

with probab

$$\overline{2^{NH(X)} \cdot 2^{NH(Y)}} = 2^{-} \; (\; ; \;)$$

Here we used

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Joint Typicality Theorem

Let $\mathbf{x}, \mathbf{y}$ be drawn from $(XY)^N$ with $P(\mathbf{x}, \mathbf{y}) = \prod_n P(x_n, y_n)$.

Joint Typicality Theorem

For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical

   $P((\mathbf{x}$

Joint Typicality Theorem

Let $\mathbf{x}, \mathbf{y}$ be drawn from $(XY)^N$ with $P(\mathbf{x}, \mathbf{y}) = \prod_n P(x_n, y_n)$.

Joint Typicality Theorem

For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical
   $P((\mathbf{x}$
2. The n
   roug

$$|J_{N\beta}| \leq 2^{N(H(X,Y)+\beta)}$$

Joint Typicality Theorem

Let $\mathbf{x}, \mathbf{y}$ be drawn from $(XY)^N$ with $P(\mathbf{x}, \mathbf{y}) = \prod_n P(x_n, y_n)$.

Joint Typicality Theorem

For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical
   $P((\mathbf{x}$
2. The n
   roug

$$|J_{N\beta}| \leq 2^{N(H(X,Y)+\beta)}$$

3. For $\mathbf{x}'$ and $\mathbf{y}'$ drawn independently from th
   marginals of $P(\mathbf{x}, \mathbf{y})$,

$$P((\mathbf{x}', \mathbf{y}') \in J_{N\beta}) \leq 2^{-N(I(X;Y)-3\beta)}$$

The Noisy-Channel Coding Theorem

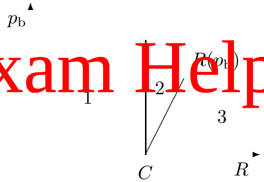
Let Q be a channel with inputs $\mathcal{A}_X$ and outputs $\mathcal{A}_Y$.

Let $C = \max_{p_X} I(X;Y)$ be the capacity of Q and $H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

The Noisy-Channel Coding Theorem

1. Any rate $R < C$ is *achievable* for Q (i. … 0, an (N, K) code with rate $K/N \geq R$ exist … $_M < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ i … (N, K) codes with rates
$$\frac{K}{N} \leq R(p_b) = \frac{C}{1 - H_2(p_b)}$$
3. For any p_b, rates greater than $R(p_b)$ are not achievable.

The Noisy-Channel Coding Theorem



Let Q be a channel with inputs $\mathcal{A}_X$ and outputs $\mathcal{A}_Y$.
Let $C = \max_{p_X} I(X;Y)$ be the capacity of Q and
$H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

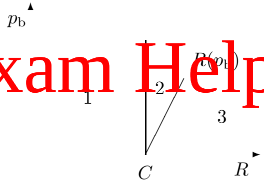
The Nois

1. Any rate $R < C$ is *achievable* for Q (i. 0, an (N, K) code with rate $K/N \geq R$ exist $_M < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ i N, K) codes with rates

$$\frac{K}{N} \leq R(p_b) = \frac{C}{1 - H_2(p_b)}$$

3. For any p_b, rates greater than $R(p_b)$ are not achievable.

The Noisy-Channel Coding Theorem



Let Q be a channel with inputs $\mathcal{A}_X$ and outputs $\mathcal{A}_Y$.
Let $C = \max_{p_X} I(X; Y)$ be the capacity of Q and
$H_2(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$.

The Nois

1. Any rate $R < C$ is *achievable* for Q (i. 0, an (N, K) code with rate $K/N \geq R$ exist $_M < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ i $N, K)$ codes with rates

$$\frac{K}{N} \leq R(p_b) = \frac{C}{1 - H_2(p_b)}$$

3. For any p_b, rates greater than $R(p_b)$ are not achievable.

Some Intuition for the NCCT

The proof of the NCCT is based on the following observations:

- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$

Some Intuition for the NCCT

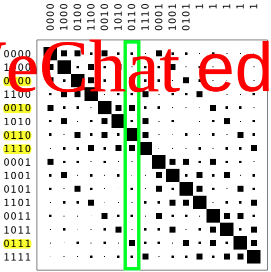
The proof of the NCCT is based on the following observations:

- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)

Some Intuition for the NCCT

The proof of the NCCT is based on the following observations:

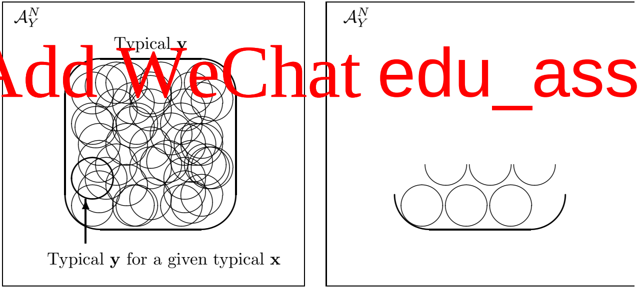
- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)
- For each $\mathbf{x}$ there are $2^{NH(Y|X)}$ typical $\mathbf{y}$ for $\mathbf{x}$



Some Intuition for the NCCT

The proof of the NCCT is based on the following observations:

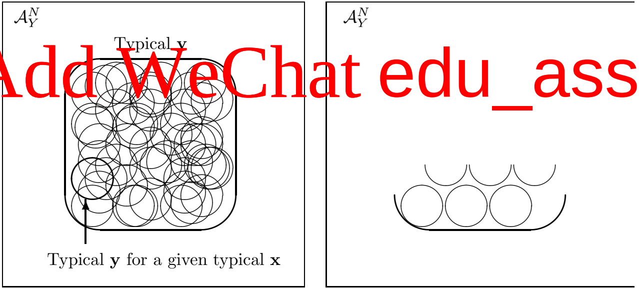
- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)
- For each $\mathbf{x}$ there are $2^{NH(Y|X)}$ typical $\mathbf{y}$ for $\mathbf{x}$
- *At most* there are $\frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{N(H(Y)-H(Y|X))} = 2^{NI(X;Y)}$ $\mathbf{x}$ with disjoint typi



Some Intuition for the NCCT

The proof of the NCCT is based on the following observations:

- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)
- For each $\mathbf{x}$ there are $2^{NH(Y|X)}$ typical $\mathbf{y}$ for $\mathbf{x}$
- *At most* there are $\frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2^{N(H(Y)-H(Y|X))} = 2^{NI(X;Y)}$ $\mathbf{x}$ with disjoint typi
- Bes s ma



Proof Sketch of NCCT Part 1

We can:

- define a family of random codes, which rely on joint typicality, and which achieve the given rate

Proof Sketch of NCCT Part 1

We can:

- define a family of random codes, which rely on joint typicality, and which achieve the given rate
- sho

Proof Sketch of NCCT Part 1

We can:

- define a family of random codes, which rely on joint typicality, and which achieve the given rate

- sho

- ded
  block error

Proof Sketch of NCCT Part 1

We can:

- define a family of random codes, which rely on joint typicality, and which achieve the given rate
- sho
- ded
  block error
- “expurgate” the above code so that it has low m
  error

This will establish that the final code achieves low maximal probability of error, while achieving the given rate!

Random Coding and Typical Set Decoding

Make **random code** $\mathcal{C}$ with rate R':

- Fix $\mathbf{p}_X$ and choose $S = 2^{NR'}$ codewords, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$ each with $P(\mathbf{x}) = \prod_n P(x_n)$

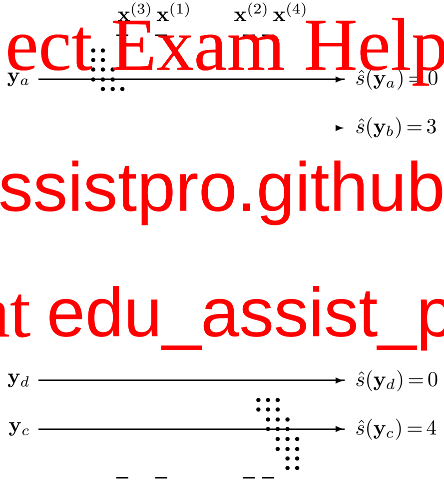
Random Coding and Typical Set Decoding

Make **random code** $\mathcal{C}$ with rate R':

- Fix $\mathbf{p}_X$ and choose $S = 2^{NR'}$ codewords, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$ each with $P(\mathbf{x}) = \prod_n P(x_n)$

Decode y

- If the
  $(\mathbf{x}^{\hat{s}}$,
  decode **y** as $\hat{s}$
- Otherwise, fail ($\hat{s} = 0$)



Random Coding and Typical Set Decoding

Make **random code** $\mathcal{C}$ with rate R':

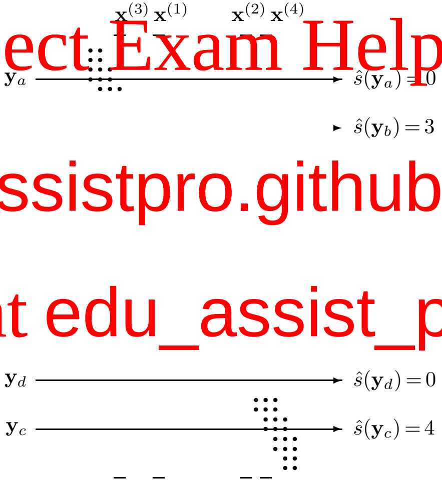
- Fix $\mathbf{p}_X$ and choose $S = 2^{NR'}$ codewords, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$ each with $P(\mathbf{x}) = \prod_n P(x_n)$

Decode y

- If the
  $(\mathbf{x}^{\hat{s}},$
  decode **y** as $\hat{s}$
- Otherwise, fail ($\hat{s} = 0$)

Errors:

- $p_B(\mathcal{C}) = P(\hat{s} \neq s|\mathcal{C})$
- $\langle p_B \rangle = \sum_{\mathcal{C}} P(\hat{s} \neq s|\mathcal{C}) P(\mathcal{C})$
- $p_{BM}(\mathcal{C}) = \max_s P(\hat{s} \neq s|s, \mathcal{C})$
  (Aim: $\exists \mathcal{C}$ s.t. $p_{BM}(\mathcal{C})$ small)



Average Error Over All Codes

Let's consider the average error over random codes:

$$\langle p_B \rangle = \quad P(\hat{s} \neq s|\mathcal{C})P(\mathcal{C})$$

A bound p es $z \in \mathcal{Z}$
with prob such that
$f(z^*)$ is smaller than the bound.

[1]If $\langle f \rangle < \delta$ but $f(z) \geq \delta$ for all z, $\langle f \rangle = \sum_z f(z)P(z) \geq \sum_z \delta P(z) = \delta$!!

Let's consider the average error over random codes:

$$\langle p_B \rangle = \quad P(\hat{s} \neq s|\mathcal{C})P(\mathcal{C})$$

A bound p … es $z \in \mathcal{Z}$ with prob … such that $f(z^*)$ is smaller than the bound.

So $\langle p_B \rangle < \delta \implies p_B(\mathcal{C}^*)$

Analogy: Suppose the average height of class is not more than 160 cm. Then one of you *must* be shorter than 160 cm.

[1] If $\langle f \rangle < \delta$ but $f(z) \geq \delta$ for all z, $\langle f \rangle = \sum_z f(z)P(z) \geq \sum_z \delta P(z) = \delta$!!

Proof Sketch of NCCT Part 1

Want to prove

> Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus

Choose s

1. Part one of the Joint Typicality Theorem says we can find an $N(\delta)$ such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.

Proof Sketch of NCCT Part 1

Want to prove

> Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus bound p_B for our random code

Choose s

1. Par $N(\delta)$
   such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.
2. Thus, the average probability of error satisfie

$$\langle p_B \rangle = \sum_{\text{atypical } (\mathbf{x},\mathbf{y})} P(\hat{s} \neq s|\cdot) \qquad \sum_{\text{typical } (\mathbf{x},\mathbf{y})} / \quad |\cdot$$

Proof Sketch of NCCT Part 1

Want to prove

> Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus bound p_B for our random code

Choose s

1. Par $N(\delta)$ such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.
2. Thus, the average probability of error satisfie

$$\langle p_B \rangle \leq \delta + \sum_{s'=2}^{2^{NR'}} 2^{-N(I(X;Y)-3\beta)}$$

Proof Sketch of NCCT Part 1

Want to prove

Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K > R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus

Choose s

1. Par $N(\delta)$
   such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.
2. Thus, the average probability of error satisfie

$$\langle p_B \rangle \leq \delta + 2^{-N(I(X;Y)-R'-3\beta)}$$

Proof Sketch of NCCT Part 1

Want to prove

> Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K > R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus

Choose s

1. Par $N(\delta)$
   such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.
2. Thus, the average probability of error satisfie

$$\langle p_B \rangle \leq \delta + 2^{-N(I(X;Y)-R'-3\beta)}$$

3. Increasing N will make $\langle p_B \rangle < 2\delta$ if $R' < I(X;Y) - 3\beta$

Proof Sketch of NCCT Part 1

Want to prove

> Any rate $R < C$ is *achievable* for Q (i.e., an (N, K) code with rate $N/K > R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Let us thus

Choose s

1. Par $N(\delta)$ such that the probability $(\mathbf{x}, \mathbf{y})$ are not jo δ.
2. Thus, the average probability of error satisfie

$$\langle p_B \rangle \leq \delta + 2^{-N(I(X;Y) - R' - 3\beta)}$$

3. Increasing N will make $\langle p_B \rangle < 2\delta$ if $R' < I(X; Y) - 3\beta$
4. Choosing maximal $P(x)$ makes required condition $R' < C - 3\beta$

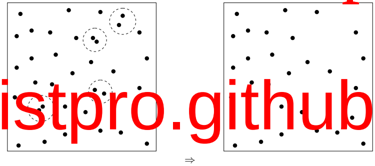
Code Expurgation

The last main "trick" is to show that if there is an (N, K) code with rate R' and $p_B(\mathcal{C}) < \delta$ we can construct a new (N, K') code $\mathcal{C}'$ with rate $R' - \frac{1}{N}$ and maximum probability of error $p_{BM}(\mathcal{C}') < 2\delta$.

Code Expurgation

The last main "trick" is to show that if there is an (N, K) code with rate R' and $p_B(\mathcal{C}) < \delta$ we can construct a new (N, K') code $\mathcal{C}'$ with rate $R' - \frac{1}{N}$ and maximum probability of error $p_{BM}(\mathcal{C}') < 2\delta$.

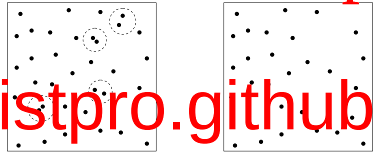
We create $\mathcal{C}'$ by **expurgating** (throwing out) half the codewords from , specifically the half with probabilit



Code Expurgation

The last main “trick” is to show that if there is an (N, K) code with rate R' and $p_B(\mathcal{C}) < \delta$ we can construct a new (N, K') code $\mathcal{C}'$ with rate $R' - \frac{1}{N}$ and maximum probability of error $p_{BM}(\mathcal{C}') < 2\delta$.

We create $\mathcal{C}'$ by **expurgating** (throwing out) half the codewords from , specifically the half with probabilit



$\Rightarrow$

Proof:

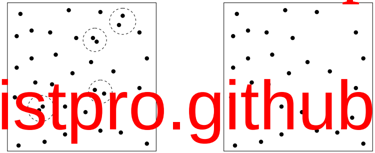
- Code $\mathcal{C}'$ has $2^{NR'}/2 = 2^{NR'-1}$ messa $' - \frac{1}{N}$
- Suppose $p_{BM}(\mathcal{C}') = \max_s P(\hat{s} \neq s|s,$ that was thrown out must have conditional probability $P(\hat{s} \neq s|s, \mathcal{C}) \geq 2\delta$
- But then

$$p_B(\mathcal{C}) = \sum_s P(\hat{s} \neq s|s, \mathcal{C})P(s) \geq \frac{1}{2}\sum_{s \notin \mathcal{C}'} 2\delta + \frac{1}{2}\sum_{s \in \mathcal{C}'} P(\hat{s} \neq s|s, \mathcal{C}) \geq \delta$$

Code Expurgation

The last main “trick” is to show that if there is an (N, K) code with rate R' and $p_B(\mathcal{C}) < \delta$ we can construct a new (N, K') code $\mathcal{C}'$ with rate $R' - \frac{1}{N}$ and maximum probability of error $p_{BM}(\mathcal{C}') < 2\delta$.

We create $\mathcal{C}'$ by **expurgating** (throwing out) half the codewords from , specifically the half with probabilit



$\Rightarrow$

Proof:

- Code $\mathcal{C}'$ has $2^{NR'}/2 = 2^{NR'-1}$ messa $' - \frac{1}{N}$.
- Suppose $p_{BM}(\mathcal{C}') = \max_s P(\hat{s} \neq s|s,$ that was thrown out must have conditional probability $P(\hat{s} \neq s|s, \mathcal{C}) \geq 2\delta$
- But then

$$p_B(\mathcal{C}) = \sum_s P(\hat{s} \neq s|s, \mathcal{C})P(s) \geq \frac{1}{2}\sum_{s \notin \mathcal{C}'} 2\delta + \frac{1}{2}\sum_{s \in \mathcal{C}'} P(\hat{s} \neq s|s, \mathcal{C}) \geq \delta$$

Wrapping It All Up

From the previous slide, $\langle p_B \rangle < 2\delta \implies$ some $\mathcal{C}'$ such that $p_{BM}(\mathcal{C}') < 4\delta$ with rate $\frac{1}{}$

Setting

NCCT Part 1: Comments

NCCT shows the existence of good codes; actually constructing practical codes is another matter

In principl

- Ho
  the t

Over the past few decades, some codes (e.g. Turb
shown to achieve rate close to the Shannon capacit

- Beyond the scope of this course!

NCCT Converse: Comments

One can in fact make a stronger statement about

$$p_{B,\text{avg}} = \frac{1}{2^K} \quad P(\mathbf{s}_{\text{out}} \neq \mathbf{s}_{\text{in}} \mid \mathbf{s}_{\text{in}}),$$

the proba

We have:

$$p_{B,\text{avg}} \geq 1 - O(e$$

Thus, if $R > C$, the probability of block error shoo ases!

- We have a “phase transition” around C between perfectly reliable and perfectly unreliable communication!

Theory and Practice

The difference between theory and practice is that, in theory, there is no difference between theory and practice but, in practice, there is.

— Jan L. A. van de Snepscheut

Theory and Practice

The difference between theory and practice is that, in theory, there is no difference between theory and practice but, in practice, there is.

— Jan L. A. van de Snepscheut

Theory v

- The
  cha
- However, the theorem is non-constructive: ... ow to create *practical* codes for a given noisy cha

- The construction of practical codes that achieve rates up to the capacity for general channels is ongoing research

Types of Codes

When we talk about types of codes we will be referring to schemes for creating (N, K) codes for any size N. MacKay makes the following distinctions:

- **Bad**: Cannot achieve arbitrarily small error, or only achieve it if the rate

  p_{BM}

Types of Codes

When we talk about types of codes we will be referring to schemes for creating (N, K) codes for any size N. MacKay makes the following distinctions:

- **Bad**: Cannot achieve arbitrarily small error, or only achieve it if the rate

  p_{BM}

- **Go** strictly less than the channel capacity (i.e, for any ϵ a good coding scheme can make a code with K/N)

Types of Codes

When we talk about types of codes we will be referring to schemes for creating (N, K) codes for any size N. MacKay makes the following distinctions:

- **Bad**: Cannot achieve arbitrarily small error, or only achieve it if the rate
  p_{BM}

- **Go**
  strictly less than the channel capacity (i.e, for any ϵ a good coding scheme can make a code with K/N)

- **Very Good**: Can achieve arbitrarily small er
  channel capacity (i.e., for any $\epsilon > 0$ a v
  make a code with $K/N = C$ and $p_{BM} < \epsilon$)

Types of Codes

When we talk about types of codes we will be referring to schemes for creating (N, K) codes for any size N. MacKay makes the following distinctions:

- **Bad**: Cannot achieve arbitrarily small error, or only achieve it if the rate
  p_{BM}
- **Go**

  strictly less than the channel capacity (i.e, for any ϵ a good coding scheme can make a code with K/N)

- **Very Good**: Can achieve arbitrarily small er
  channel capacity (i.e., for any $\epsilon > 0$ a v
  make a code with $K/N = C$ and $p_{BM} < \epsilon$)
- **Practical**: Can be coded and decoded in time that is polynomial in the block length N.

Random Codes

During the discussion of the Noisy-Channel Coding Theorem we saw how to construct very good **random codes** via typical set decoding

Properties:

- Ver ... rror
- Co
- Not Practical:
  - The 2^K codewords have no structure an
  - Typical set decoding is expensive

Linear Codes

(N, K) Block Code

An (N, K) **block code** is a list of $S = 2^K$ codewords $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}\}$, each of length N. A message $s \in \{1, 2, \ldots, 2^K\}$ is encoded as $\mathbf{x}^{(s)}$.

Linear Codes

(N, K) Block Code

An (N, K) **block code** is a list of $S = 2^K$ codewords $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}\}$, each of length N. A message $s \in \{1, 2, \ldots, 2^K\}$ is encoded as $\mathbf{x}^{(s)}$.

Linear (

A **linear** first represe ed via multiplication by an $N \times K$ binary matrix $\mathbf{G}$ lo 2.

Here linear means all $S = 2^K$ messages ca different combinations of the K codewords -bit string with single 1 in position i.

Linear Codes

(N, K) Block Code

An (N, K) **block code** is a list of $S = 2^K$ codewords $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}\}$, each of length N. A message $s \in \{1, 2, \ldots, 2^K\}$ is encoded as $\mathbf{x}^{(s)}$.

Linear (

A **linear** … first … represe… ed via multiplication by an $N \times K$ binary matrix $\mathbf{G}$ … lo 2.

Here linear means all $S = 2^K$ messages ca… different combinations of the K codewords … -bit string with single 1 in position i.

Example: Suppose $(N, K) = (7, 4)$. To send $s = 3$, first create $\mathbf{s} = 0011$ and send $\mathbf{t} = \mathbf{G}^\top \mathbf{s} = \mathbf{G}^\top(\mathbf{e}_0 + \mathbf{e}_1) = \mathbf{G}^\top \mathbf{e}_0 + \mathbf{G}^\top \mathbf{e}_1 = \mathbf{t}_0 + \mathbf{t}_1$ where $\mathbf{e}_0 = 0001$ and $\mathbf{e}_1 = 0010$.

Types of Linear Code

Many commonly used codes are linear:

- Repetition Codes: e.g., 0 $\rightarrow$ 000 ; 1 $\rightarrow$ 111
- Convolution Codes: Linear coding plus bit shifts
- Concatenation Codes: Two or more levels of error correction
- Ha
- Lo

Types of Linear Code

Many commonly used codes are linear:

- Repetition Codes: e.g., $0 \rightarrow 000$; $1 \rightarrow 111$
- Convolution Codes: Linear coding plus bit shifts
- Concatenation Codes: Two or more levels of error correction
- Ha
- Lo

An NCCT

replacing “there exists a code”) but the proof is still non

Types of Linear Code

Many commonly used codes are linear:

- Repetition Codes: e.g., $0 \rightarrow 000$; $1 \rightarrow 111$
- Convolution Codes: Linear coding plus bit shifts

- Concatenation Codes: Two or more levels of error correction
- Ha
- Lo

An NCCT
replacing "there exists a code") but the proof is still non

Practical linear codes:

- Use very large block sizes N
- Based on semi-random code constructions
- Apply probabilistic decoding techniques
- Used in wireless and satellite communication

Linear Codes: Examples

(7,4) Hamming Code

(6,3) Repetition Code

1 0 0 0

1 0 0

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

For $\mathbf{s} = 0011$,

$$\mathbf{G}^\top \mathbf{s} (\bmod 2) = [0\ 0\ 1\ 1\ 1\ 0\ 0]^\top$$

For

$$\mathbf{G}^\top \mathbf{s} (\bmod 2) = [0\ 1\ 0\ 0\ 1\ 0]^\top$$

Decoding

We can construct codes with a relatively simple encoding but how do we decode them? That is, given the input distribution and channel model Q how do we find the posterior distribution over $\mathbf{x}$ given we received $\mathbf{y}$?

Decoding

We can construct codes with a relatively simple encoding but how do we decode them? That is, given the input distribution and channel model Q how do we find the posterior distribution over $\mathbf{x}$ given we received $\mathbf{y}$?

Simple? J

$$\overline{\quad}$$
$$\mathbf{x}' \in \mathcal{C} \; P(\mathbf{y}\ \mathbf{x})P(\mathbf{x})$$

But:

- the number of codes $\mathbf{x} \in \mathcal{C}$ is 2^K so, n
- linear codes provide structure that the above method doesn't exploit

Summary and Reading

Main Points:

- Joint Typicality and the Joint Typicality Theorem

- The (Longer) Noisy Channel Coding Theorem
- Pro

  - Average Error Over Random Codes

  - Code Expurgation

Reading:

- MacKay §9.7, §10.1-§10.5