COMP2610/6261 - Information Theory

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

Mark Reid and Aditya Menon

Research School of Computer Science
The Australian National University

October 8th, 2014

The Noisy-Channel Coding Theorem
Informal Statement

Noisy-Channel Coding Theorem (Informal)

If Q is a channel with capacity C then the rate R is *achievable* **if and only if** $R \leq C$, that is, the rate is no greater than the channel capacity.

The Noisy-Channel Coding Theorem

Informal Statement

Noisy-Channel Coding Theorem (Informal)

If Q is a channel with capacity C then the rate R is *achievable* **if and only if** $R \leq C$, that is, the rate is no greater than the channel capacity.

The Noisy-Channel Coding Theorem (Formal)

1. Any rate $R < C$ is *achievable* for Q (i.e., 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

Informal Statement

Noisy-Channel Coding Theorem (Informal)

If Q is a channel with capacity C then the rate R is *achievable* **if and only if** $R \leq C$, that is, the rate is no greater than the channel capacity.

The Noisy-Channel Coding Theorem (Formal)

1. Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ is acceptable, (N, K) codes exists with rates

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

Noisy-Channel Coding Theorem (Informal)

If Q is a channel with capacity C then the rate R is *achievable* **if and only if** $R \leq C$, that is, the rate is no greater than the channel capacity.

The Noisy-Channel Coding Theorem (Formal)

1. Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ is acceptable, (N, K) codes exists 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.

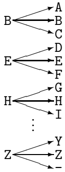
The Noisy-Channel Coding Theorem
For the Noisy Typewriter

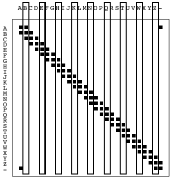
NCCT

Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)

For noisy typewriter Q:

- The capacity is $C = \log_2 9$
- For any $\epsilon > 0$ and $R < C$ we can choose $N = 1$ ...
- ...and code messages using $\mathcal{C} = \{\mathtt{B}, \mathtt{E}, \ldots, \mathtt{Z}\}$





The Noisy-Channel Coding Theorem

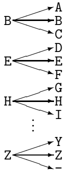
For the Noisy Typewriter

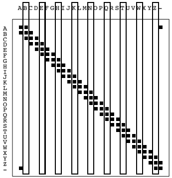
NCCT

Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)

For noisy typewriter Q:

- The capacity is $C = \log_2 9$
- For any $\epsilon > 0$ and $R < C$ we can choose $N = 1$ ...
- ... and code messages using $\mathcal{C} = \{\mathtt{B}, \mathtt{E}, \ldots, \mathtt{Z}\}$





Since $|\mathcal{C}| = 9$ we have $K = \log_2 9$ so $K/N = \log_2 9 \geq R$ for any $R < C$, and $\mathcal{C}$ has zero error so $p_{BM} = 0 < \epsilon$

Joint Typicality

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

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

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 $H(Z)$

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

Joint Typicality

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

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

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

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

Joint Typicality

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

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

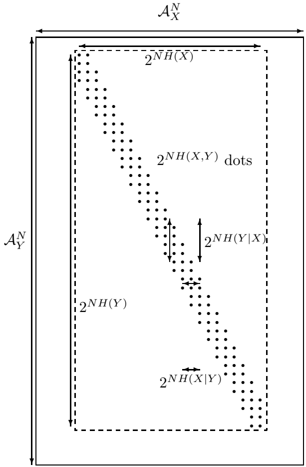
The **jointly typical set** of all such pairs is denoted $J_{N\beta}$.

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

$\mathbf{x}$ 1111111111000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

$\mathbf{y}$ 0011111111000000000000000000000000000000000000000000000000000000000000000000000000111111111111111111

Joint Typicality



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

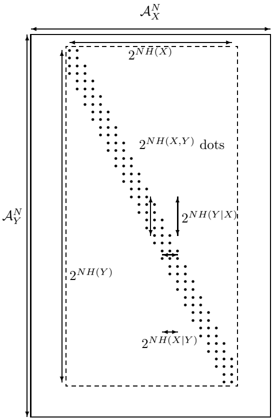
Joint Typicality Theorem

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

Joint Typicality Theorem

For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical
   $P((\mathbf{x}, \mathbf{y}) \in J_{N\beta}) \to 1$ as $N \to \infty$



Joint Typicality Theorem

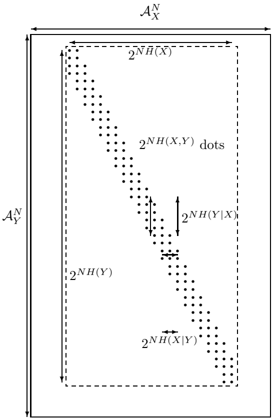
Let $\mathbf{x}, \mathbf{y}$ be drawn from $(XY)^N$ with distribution $P(\mathbf{x}, \mathbf{y}) = \prod_n P(x_n, y_n)$.

Joint Typicality Theorem

For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical $P((\mathbf{x}, \mathbf{y}) \in J_{N\beta}) \to 1$ as $N \to \infty$
2. The number of jointly typical sequences is roughly $2^{NH(X,Y)}$:

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



Joint Typicality Theorem

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

Joint Typicality Theorem

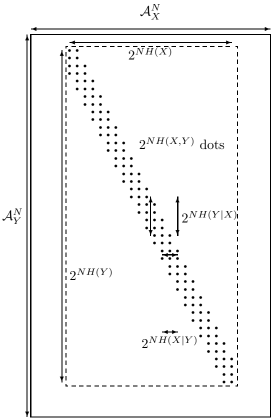
For all tolerances $\beta > 0$

1. Almost every pair is eventually jointly typical $P((\mathbf{x}, \mathbf{y}) \in J_{N\beta}) \to 1$ as $N \to \infty$
2. The number of jointly typical sequences is roughly $2^{NH(X,Y)}$:

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

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

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



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

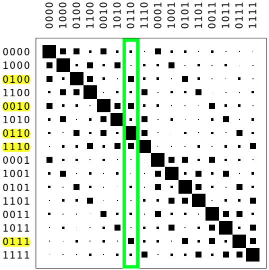
The proof of the NCCT is based on the following observations:

- 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:

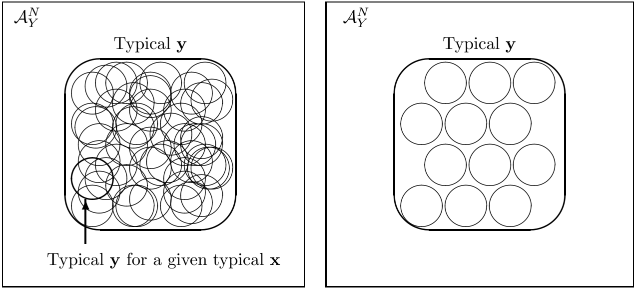
- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)
- For each $\mathbf{x}$ there are $2^{NH(Y|X)}$ typical $\mathbf{y}$ for $\mathbf{x}$



Some Intuition for the NCCT

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

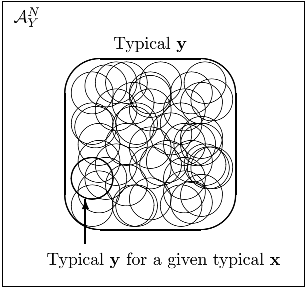
- Each choice of input distribution $\mathbf{p}_X$ induces an output distibution $\mathbf{p}_Y$
- There are $2^{NH(Y)}$ typical $\mathbf{y}$ (i.e., with prob. per symbol $\approx H(Y)$)
- For each $\mathbf{x}$ there are $2^{NH(Y|X)}$ typical $\mathbf{y}$ for $\mathbf{x}$
- *At most* there are $\frac{2^{NH(Y)}}{2^{NH(Y|X)}}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

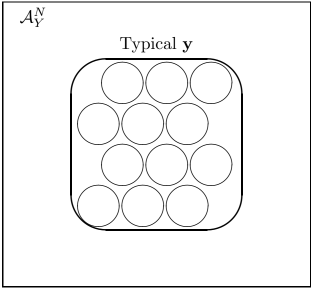


Some Intuition for the NCCT

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 typical $\mathbf{y}$. Coding with these $\mathbf{x}$ minimises error
- Best rate K/N achieved when number of such $\mathbf{x}$ (i.e., 2^K) is maximised: $2^K \leq \max_{\mathbf{p}_X} 2^{NI(X;Y)} = 2^{N \max_{\mathbf{p}_X} I(X;Y)} = 2^{NC}$





The Noisy-Channel Coding Theorem

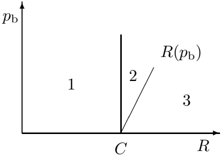
Let Q be a channel with inputs $\mathcal{A}_X$ and outputs $\mathcal{A}_Y$.
Let $C = \max_{\mathbf{p}_X} I(X;Y)$ be the capacity of Q and
$H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

The Noisy-Channel Coding Theorem

1. Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ is acceptable, (N, K) codes exists 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 $C = \max_{\mathbf{p}_X} I(X; Y)$ be the capacity of Q and
$H_2(p) = -p \log_2 p - (1-p)\log_2(1-p)$.

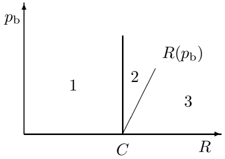


The Noisy-Channel Coding Theorem

1. Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)
2. If probability of bit error $p_b := p_B/K$ is acceptable, (N, K) codes exists 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 $C = \max_{\mathbf{p}_X} I(X;Y)$ be the capacity of Q and
$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$.



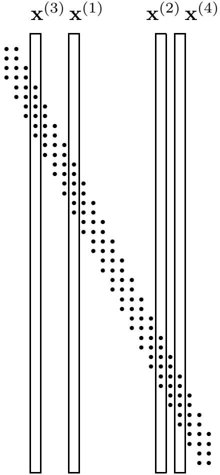
The Noisy-Channel Coding Theorem

1. Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)
2. If probability of bit error $p_b := p_B / K$ is acceptable, (N, K) codes exists 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.

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 $P(\mathbf{x}) = \prod_n P(x_n)$



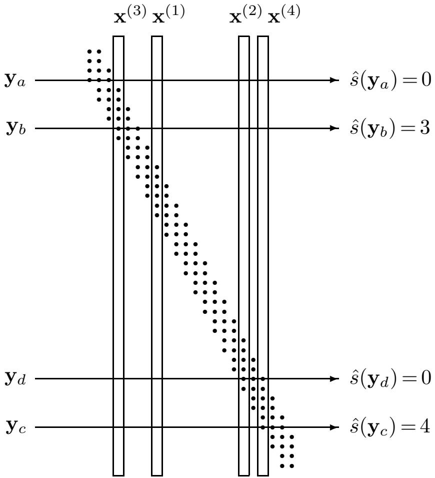
Random Coding and Typical Set Decoding

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 via typical sets:

- If there is *exactly one* $\hat{s}$ so that $(\mathbf{x}^{\hat{s}}, \mathbf{y})$ are jointly typical then decode $\mathbf{y}$ as $\hat{s}$
- Otherwise, fail ($\hat{s} = 0$)



Random Coding and Typical Set Decoding

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

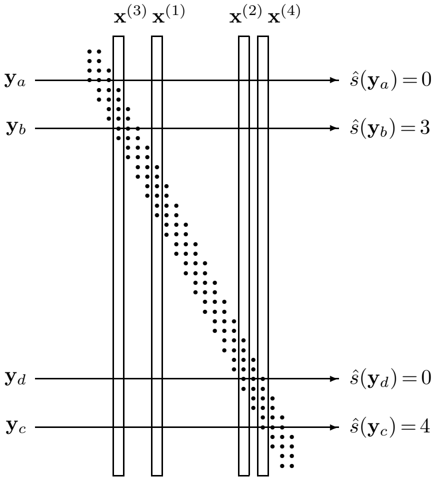
- Fix $\mathbf{p}_X$ and choose $S = 2^{NR'}$ codewords, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}$, each with $P(\mathbf{x}) = \prod_n P(x_n)$

Decode y via typical sets:

- If there is *exactly one* $\hat{s}$ so that $(\mathbf{x}^{\hat{s}}, \mathbf{y})$ are jointly typical then decode $\mathbf{y}$ as $\hat{s}$
- Otherwise, fail ($\hat{s} = 0$)

Errors:

- $p_B(\mathcal{C}) = P(\hat{s} \neq s|\mathcal{C})$
- $p_B = \sum_{\mathcal{C}} P(\hat{s} \neq s|\mathcal{C})P(\mathcal{C})$
- $p_{BM}(\mathcal{C}) = \max_s P(\hat{s} \neq s|s, \mathcal{C})$
  (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:

$$p_B = \sum_{\mathcal{C}} P(\hat{s} \neq s|\mathcal{C})P(\mathcal{C})$$

A bound on the average $\bar{f}$ of some function f of random variables $z \in \mathcal{Z}$ with probabilities $P(z)$ *guarantees* there is at least one $z^* \in \mathcal{Z}$ such that $f(z^*)$ is smaller than the bound.[1]

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

Average Error Over All Codes

Let's consider the average error over random codes:

$$p_B = \sum_{\mathcal{C}} P(\hat{s} \neq s | \mathcal{C}) P(\mathcal{C})$$

A bound on the average $\bar{f}$ of some function f of random variables $z \in \mathcal{Z}$ with probabilities $P(z)$ *guarantees* there is at least one $z^* \in \mathcal{Z}$ such that $f(z^*)$ is smaller than the bound.[1]

$$\text{So } p_B < \delta \implies p_B(\mathcal{C}^*) < \delta \text{ for some } \mathcal{C}^*.$$

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 $\bar{f} < \delta$ but $f(z) \geq \delta$ for all z, $\bar{f} = \sum_z f(z)P(z) \geq \sum_z \delta P(z) = \delta$!!

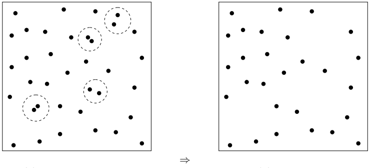
Code Expurgation

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 $p_{BM}(\mathcal{C}') < 2\delta$.

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 $p_{BM}(\mathcal{C}') < 2\delta$.

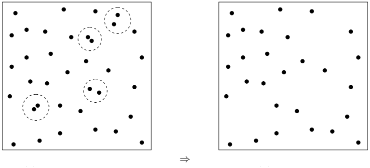
We create $\mathcal{C}'$ by **expurgating** (throwing out) half the codewords from $\mathcal{C}$, specifically the half with the largest *conditional* probability of error.



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 $\mathcal{C}$, specifically the half with the largest *conditional* probability of error.



Proof:

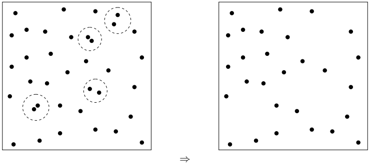
- Code $\mathcal{C}'$ has $2^{NR}/2 = 2^{NR-1}$ messages, so rate of $K'/N = R - \frac{1}{N}$.
- Suppose $p_{BM}(\mathcal{C}') = \max_s P(\hat{s} \neq s | s, \mathcal{C}') \geq 2\delta$, then every $s \in \mathcal{C}$ that was thrown out must have conditional probability $P(\hat{s} \neq s | s, \mathcal{C}) \geq 2\delta$
- 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 $\mathcal{C}$, specifically the half with the largest *conditional* probability of error.



Proof:

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

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

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)

Choose some $\delta > 0$

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 jointly typical is less than 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)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

$$p_B = \sum_{\text{atypical } (\mathbf{x},\mathbf{y})} P(\hat{s} \neq s|\cdot) + \sum_{\text{typical } (\mathbf{x},\mathbf{y})} P(\hat{s} \neq s|\cdot)$$

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)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

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

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 $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

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

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 $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)
$$p_B \leq \delta + 2^{-N(I(X;Y)-R'-3\beta)}$$
3. Increasing N will make $p_B < 2\delta$ 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 $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

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

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

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 $N/K \geq R$ exists with max. block error $p_{BM} < \epsilon$ for any tolerance ϵ)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

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

3. Increasing N will make $p_B < 2\delta$ if $R' < I(X;Y) - 3\beta$
4. Choosing maximal $P(x)$ makes required condition $R' < C - 3\beta$
5. $p_B < 2\delta \implies$ a $\mathcal{C}'$ such that $p_{BM}(\mathcal{C}') < 4\delta$ with rate $R' - \frac{1}{N}$

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)

Choose some $\delta > 0$

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 jointly typical is less than δ.
2. Thus, the average probability of error satisfies (by Part 3 of JCT)

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

3. Increasing N will make $p_B < 2\delta$ if $R' < I(X; Y) - 3\beta$
4. Choosing maximal $P(x)$ makes required condition $R' < C - 3\beta$
5. $p_B < 2\delta \implies$ a $\mathcal{C}'$ such that $p_{BM}(\mathcal{C}') < 4\delta$ with rate $R' - \frac{1}{N}$
6. Setting $R' = (R + C)/2, \delta = \epsilon/4, \beta < (C - R')/3$ gives the result.

Summary and Reading

Main Points:

- Joint Typicality and the Joint Typicality Theorem
- The (Longer) Noisy Channel Coding Theorem
- Proof Ideas
  - Random Coding & Typical Set Decoding
  - Average Error Over Random Codes
  - Code Expurgation

Reading:

- MacKay §9.7, §10.1-§10.5