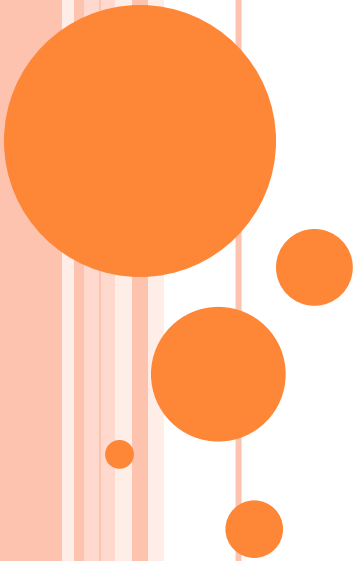
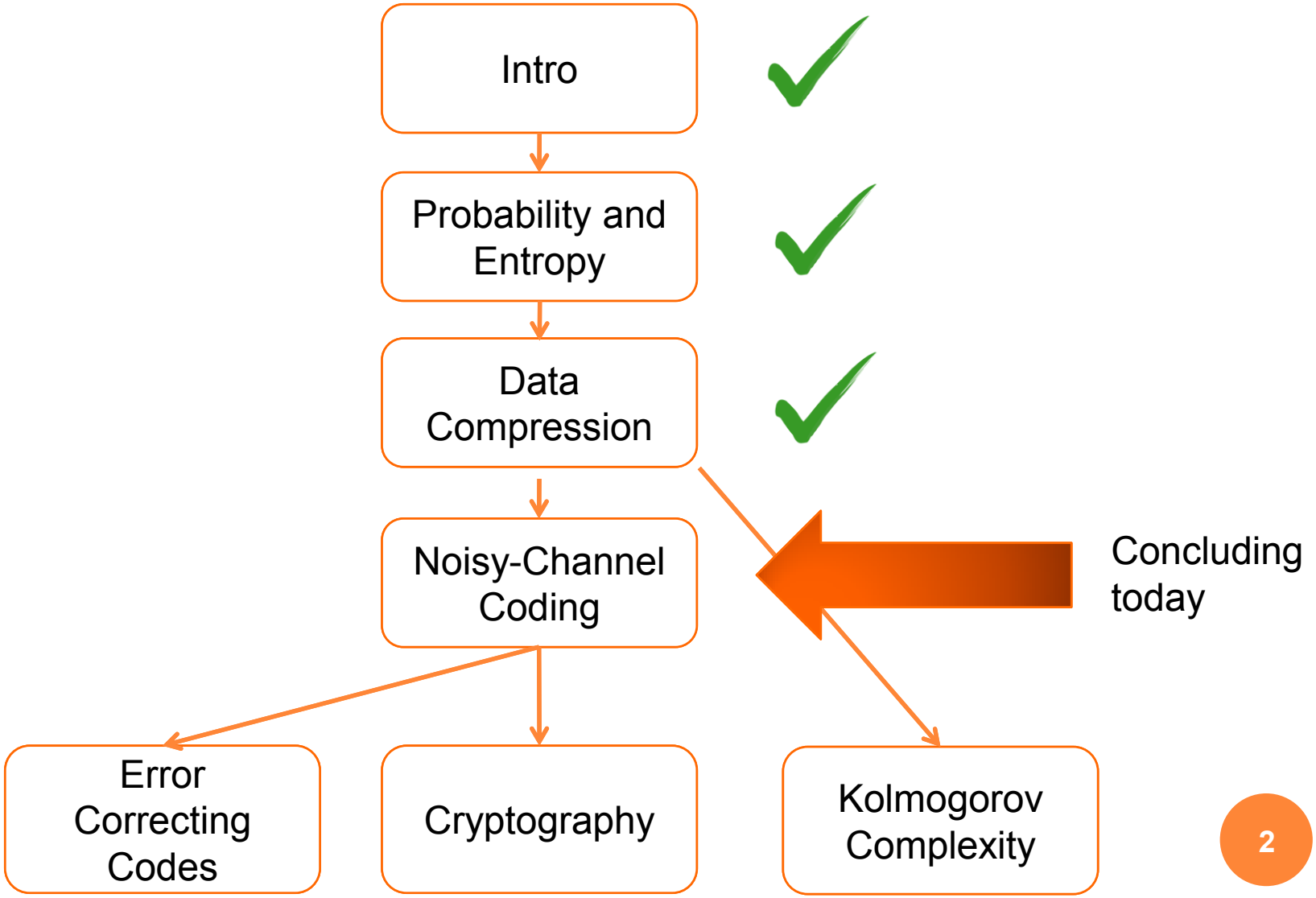


CS3236 Introduction to Information theory

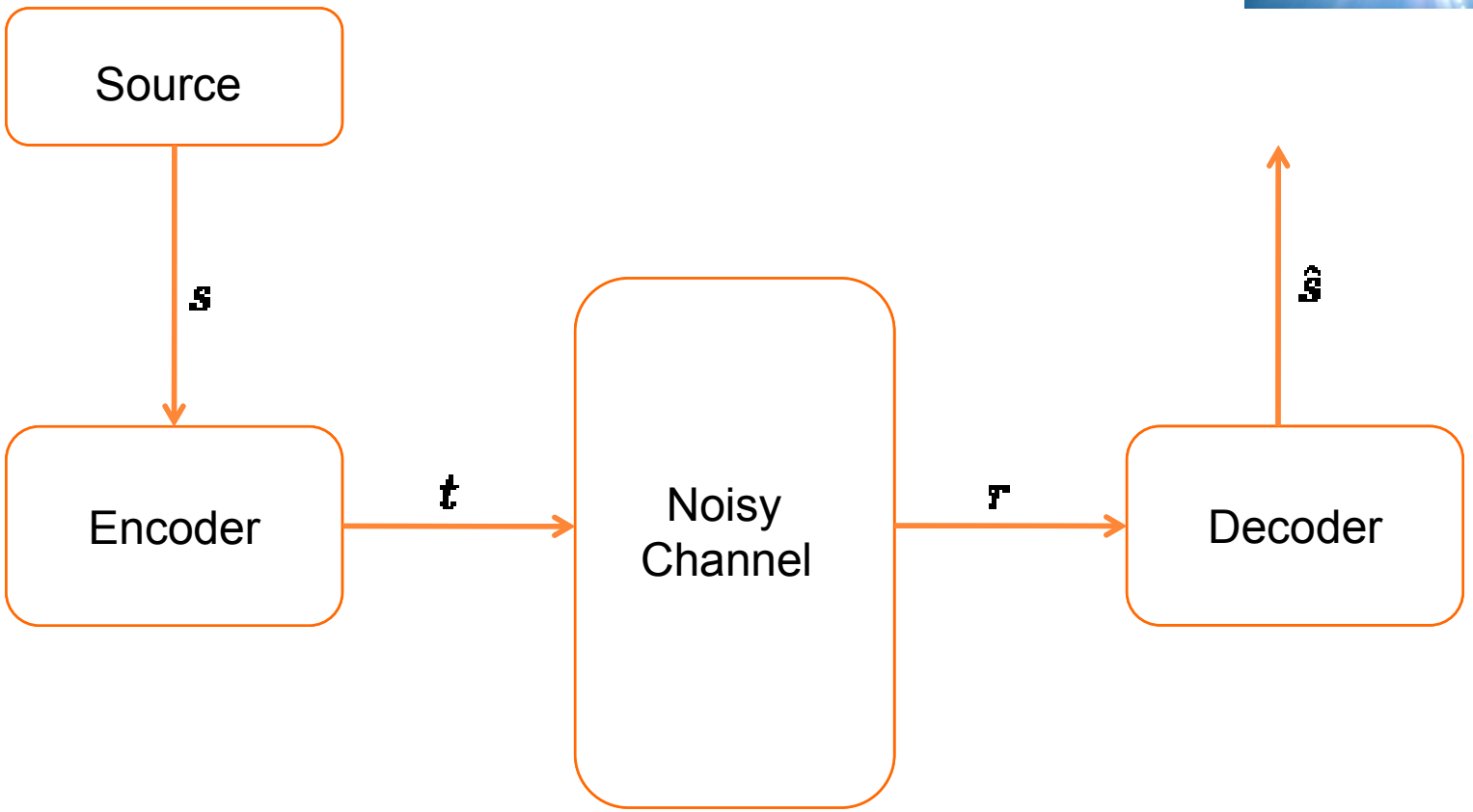
Lecture 7: Noisy-channel coding theorem

Course given by Pierre Senellart
Material by Stephanie Wehner, with additions by P. Senellart

Where do we go from here?

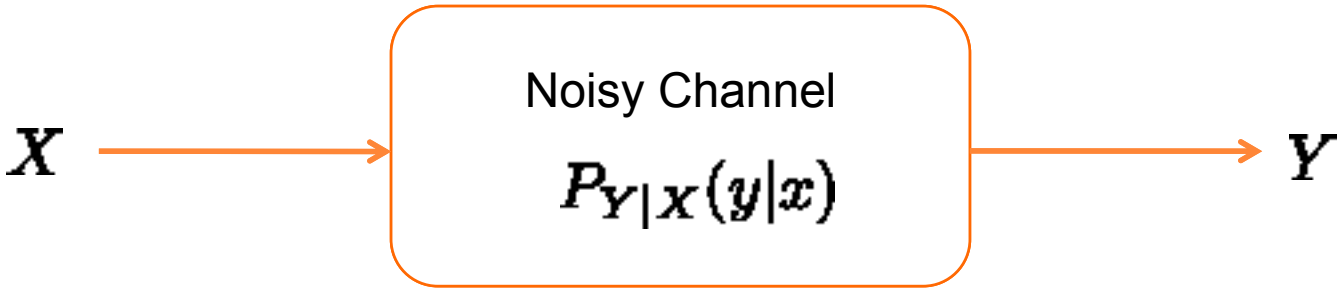


Sending information



Goal: Construct an encoder and decoder such that we can recover the information $\hat{s} = s$ with very low error

How much information does the output give about the input?



- Intuition: mutual information measures how much our uncertainty about X is reduced by learning Y

$$I(X;Y) = H(X) - H(X|Y)$$

$$I(X;Y) = H(Y) - H(Y|X)$$

Capacity of a channel

- Shannon proved that the capacity of a channel $\mathcal{N}$ given by

$$C(\mathcal{N}) = \max_{P_X} I(X;Y)$$

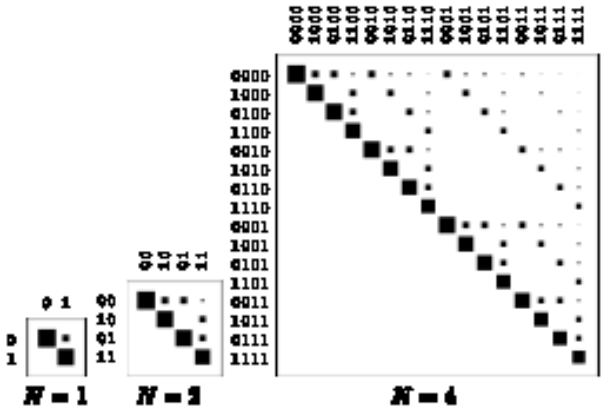
- Determines how well we can send information.

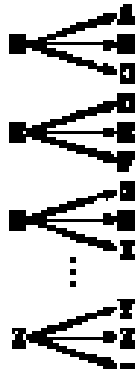
Shannon's noisy channel coding theorem (Part 1: Achievability)

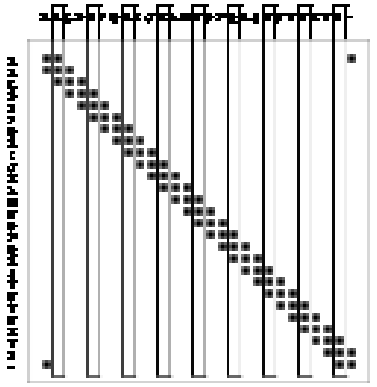
- Associated with every discrete memoryless channel, there is a non-negative number *C* (the capacity) such that
  - For any error $\epsilon > 0$ and $R \leq C$ for large enough *N*, there exists a block code of length *N* and rate *R*, an a decoding algorithm, such that the maximum probability of block error is $< \epsilon$

Essential idea

- For large N, the diagrams for the extended channel start to look a lot like the noisy-typewriter!
- Restricting to typical inputs and outputs allows us to do the same general channels







Let's first gather an essential tool.. Joint typicality!

Typical inputs: x typical for P(x)

- Remember the lecture on the source coding theorem!
- We say that some iid input $x_1, \ldots, x_N$ with probability $P(x) = P(x_1) \cdot \ldots \cdot P(x_N)$ is typical if

$$\left| \frac{1}{N} \log \frac{1}{P(x)} - H(X) \right| < \beta$$

- Size of the typical set was roughly $2^{NH(X)}$
- All elements of the typical set occur with roughly the same probability

Typical outputs: Y typical for P(Y)

- For the outputs, we can also define the typical set!
- For iid input the output distribution is also iid

$$\left| \frac{1}{N} \log \frac{1}{P(y)} - H(Y) \right| < \beta$$

- Size of the typical set is roughly $2^{NH(Y)}$

- All elements of the typical set occur with roughly the same probability

Joint typicality: (x,y) typical for P(x,y)

- Same idea, but now for the joint distribution

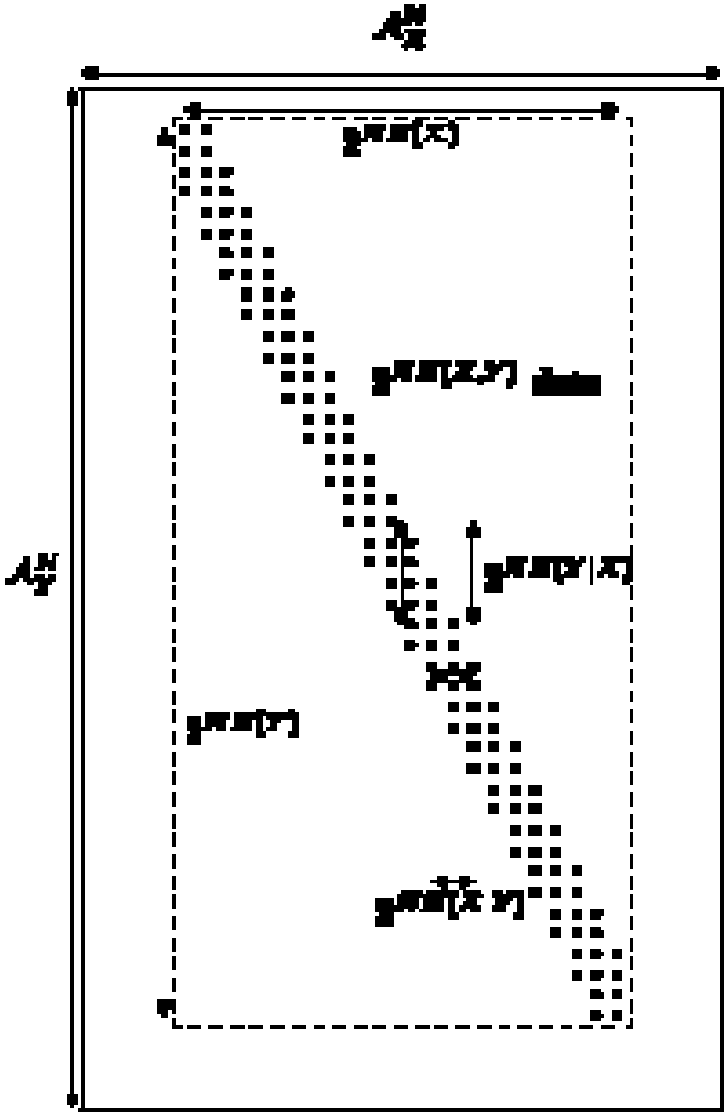
$$P(x,y) = P(x_1,y_1)\ldots p(x_N,y_N)$$

- A sequence x,y is jointly typical if
  - x is typical for P(x)
  - y is typical of P(y)
  - (x,y) is typical for P(x,y)

$$\left| \frac{1}{N} \log \frac{1}{P(x,y)} - H(X,Y) \right| < \beta$$

- Size of the typical set roughly $2^{NH(X,Y)}$

JOINTLY TYPICAL SETS - DIAGRAM



Drawing X and Y independently

- Marginals should be the same

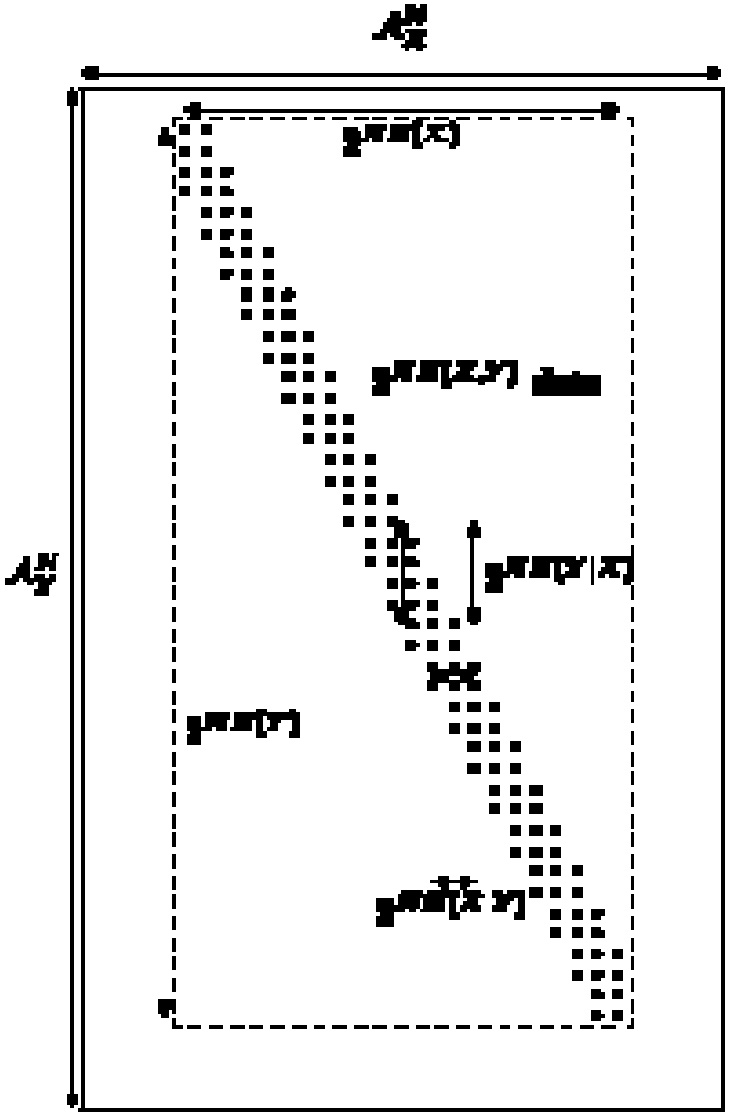
$$P(x') := \sum_y P(x', y)$$

$$P(y') := \sum_x P(x, y')$$

- Draw x' and y' from

$$P(x', y') := P(x')P(y')$$

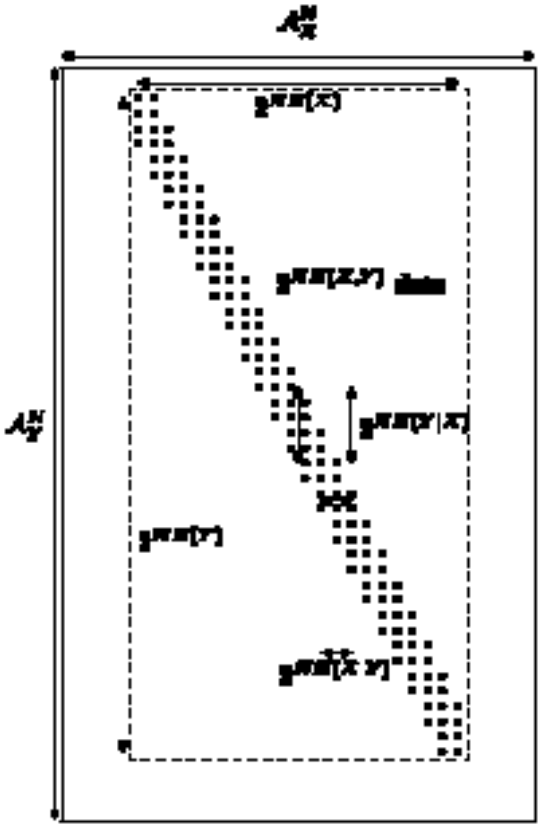
- Are x' and y' likely to be in the original jointly typical set?



Embedding in the jointly typical set

- Original jointly typical set

$$J_{N\beta} := \{(x,y) \mid \left|\tfrac{1}{N}\log\tfrac{1}{P(x)} - H(X)\right| \left|\tfrac{1}{N}\log\tfrac{1}{P(y)} - H(Y)\right| \left|\tfrac{1}{N}\log\tfrac{1}{P(x,y)} - H(X,Y)\right|\}$$

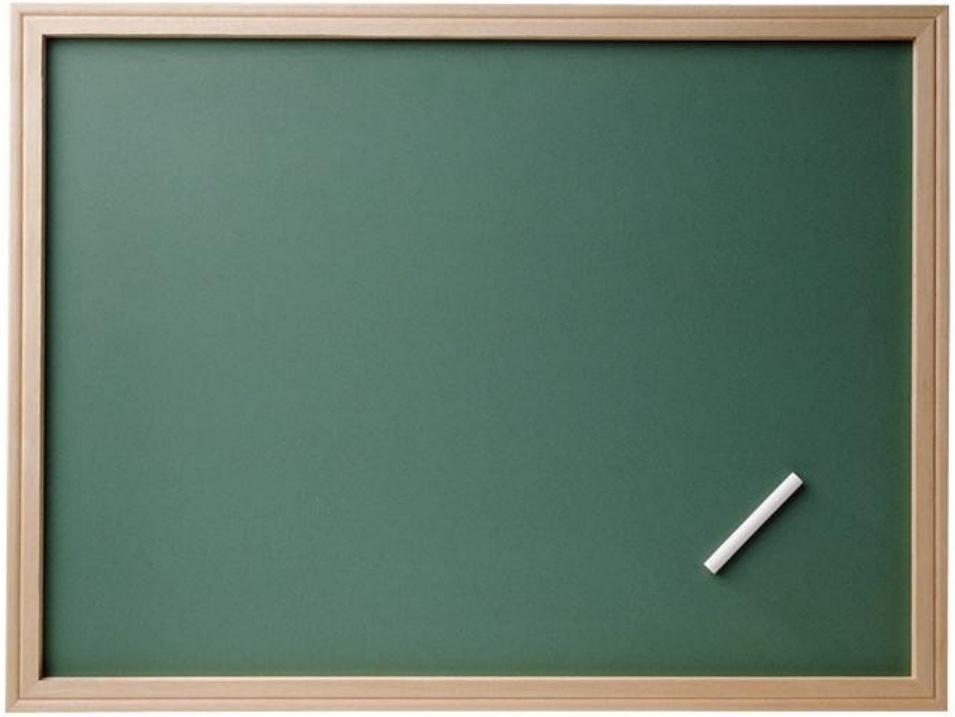


- From our picture we expect

$$P((x',y') \in J_{N\beta}) \approx \frac{2^{NH(X,Y)}}{2^{N(H(X)+H(Y))}} = 2^{-NI(X;Y)}$$

Computation

- Let's verify: $$P((x', y') \in J_{N\beta}) \approx \frac{2^{NH(X,Y)}}{2^{N(H(X)+H(Y))}} = 2^{-NI(X;Y)}$$



SUMMARY: JOINT TYPICALITY THEOREM

- Consider strings x and y with $P(x,y) = \prod_{j=1}^{N} P(x_j, y_j)$
- 1. The probability that x and y are in the typical set goes to 1 for large N
- 2. The number of jointly typical sequences obeys

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

  - Proofs same as in source coding theorem
- 3. For x' and y' drawn independently according to the marginal distributions

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

  - Just proved this!

One more trick….

A USEFUL TRICK: AVERAGING IS SOMETIMES EASIER

- Imagine we have 100 durians, and we want to determine whether at least one of them weighs 1.5 kg
- Question: in the best case, how many times do we have to use a (large) weighing scale?

A USEFUL TRICK: AVERAGING IS SOMETIMES EASIER

- If we are lucky, just once!
- Weigh them all together, and compute average weight $\frac{1}{100}W$

$\frac{1}{100}W \geq 1.5kg$ There is at least one durian of weight 1.5kg

$\frac{1}{100}W < 1.5kg$ Hmm, who knows? There might be a big durian if there are many small ones too.

How is this useful?

- We want to show that there exists a code with rate achieving the capacity
- Average over all possible codes!

Average block error for code

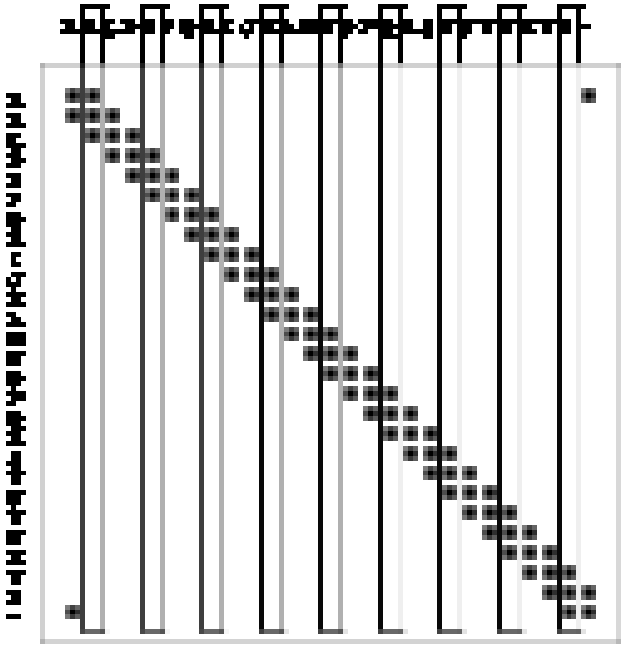
$$\mathbb{E}(P_{\text{err,Block}}) = \sum_{\text{code } C} P(C)P(\hat{s} \neq s|C)$$

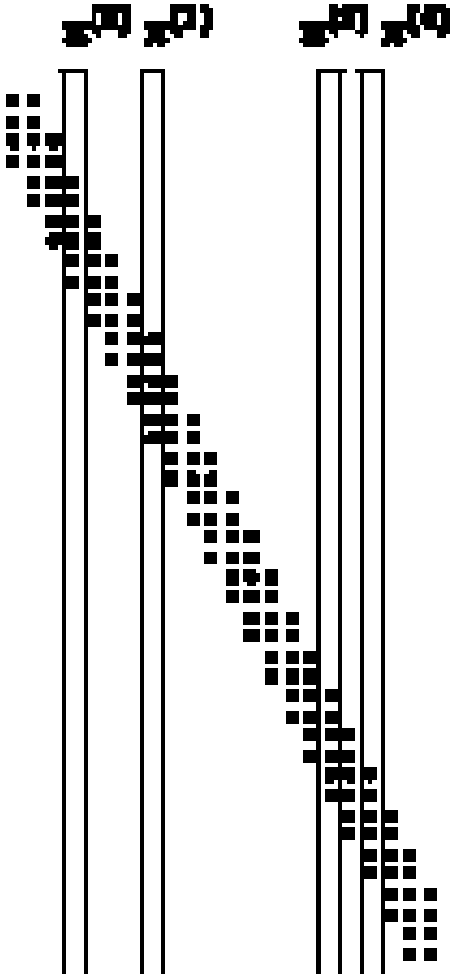
- If we can show that this average is small for codes with a rate at capacity then we know there exists at least one such code!
- … later worry about maximum block error

What does it mean to choose codes randomly?

- Generate $2^{NR'}$ codewords for a $(N, NR') = (N, K)$ code

Compare noisy-typewriter





For a random code

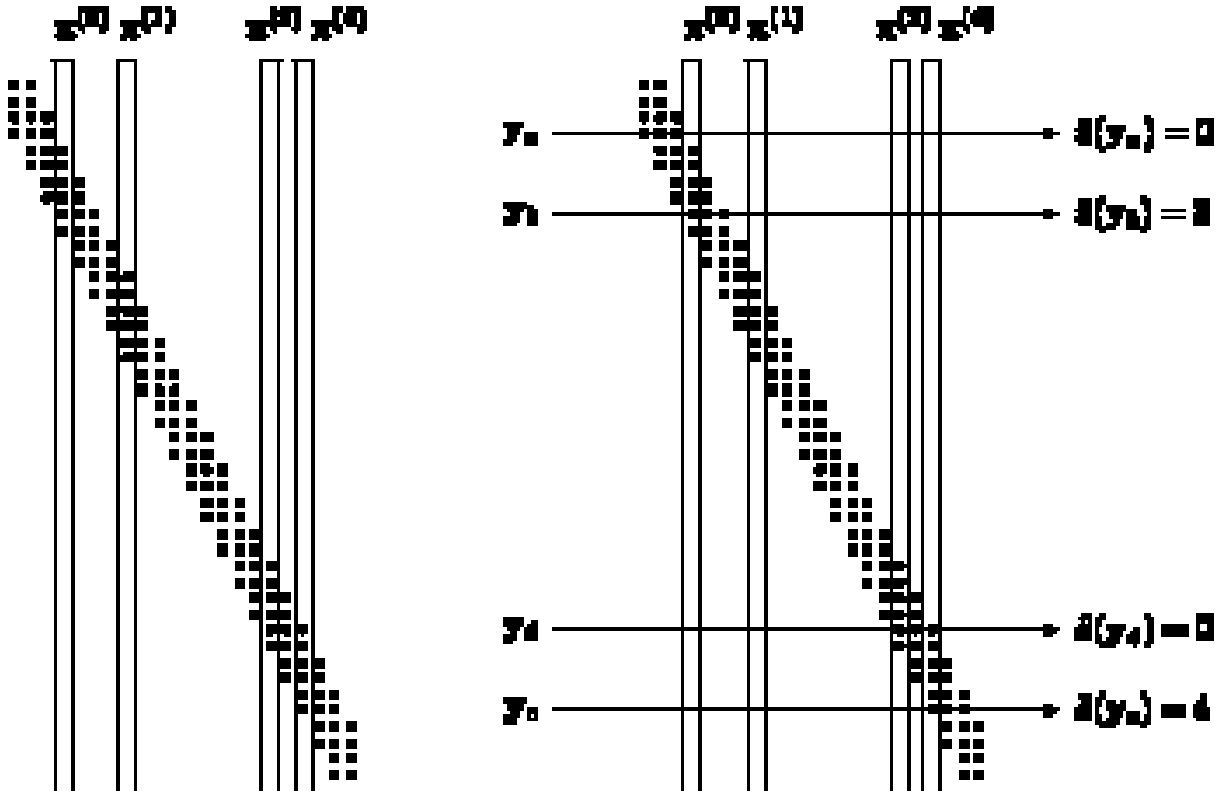
- Sender and receiver both know the code!
- A source symbol $s \in \{1, \ldots, 2^{NR'}\}$ is encoded as $x^{(s)}$
- The received signal is y with probability

$$P(y|x^{(s)}) = \prod_{j=1}^{N} P(y_j|x_j^{(s)})$$

Onto to the Proof!

OUTLINE

- Choose a random code, and random inputs s
- Typical set decoding: Given y output $\hat{s}$ such that
  - $(x^{(\hat{s})}, y)$ are jointly typical
  - There is no other s' such that $(x^{(s')}, y)$ are jointly typical
  - Otherwise output failure $\hat{s} = 0$



How well can this work?

- Two ways our typical set decoder can fail
  - There's no $\hat{s}$ such that $(x^{(\hat{s})}, y)$ is jointly typical
  - There is another s' such that $(x^{(s')}, y)$ is jointly typical

- Since all codewords play the same role, it is sufficient to show what happens for some fixed input $s = 1$

- Want to show that the probability both failure events happen is small

Part 1

- Want to show that the probability that there is no jointly typical element can be made arbitrarily small

- Joint typicality theorem, part 1:

$$P((x^{(1)}, y) \in J_{N\beta}) \geq 1 - \delta$$

- But then there exists at least one jointly typical element (with high probability)

Part 2

- Want to show that there is no other s' such that $(x^{(s')}, y)$ is jointly typical

- By the joint typicality theorem for any fixed $s' \neq 1$

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

- Since there are $2^{NR'} - 1$ values for $s' \neq 1$

$$\begin{aligned}\mathbb{E}(p_{\text{err,Block}}) &\leq \delta + \sum_{s'=2}^{2^{NR'}} 2^{-N(I(X;Y)-3\beta)} \\ &\leq \delta + 2^{-N(I(X;Y)-R'-3\beta)}\end{aligned}$$

Wrapping it up

- We know the average block error obeys

$$\mathbb{E}(p_{\text{err,Block}}) \leq \delta + 2^{-N(I(X;Y)-R'-3\beta)}$$

- Hence for $R' < I(X;Y) - 3\beta$ we have for large enough N

$$\mathbb{E}(p_{\text{err,Block}}) \leq 2\delta$$

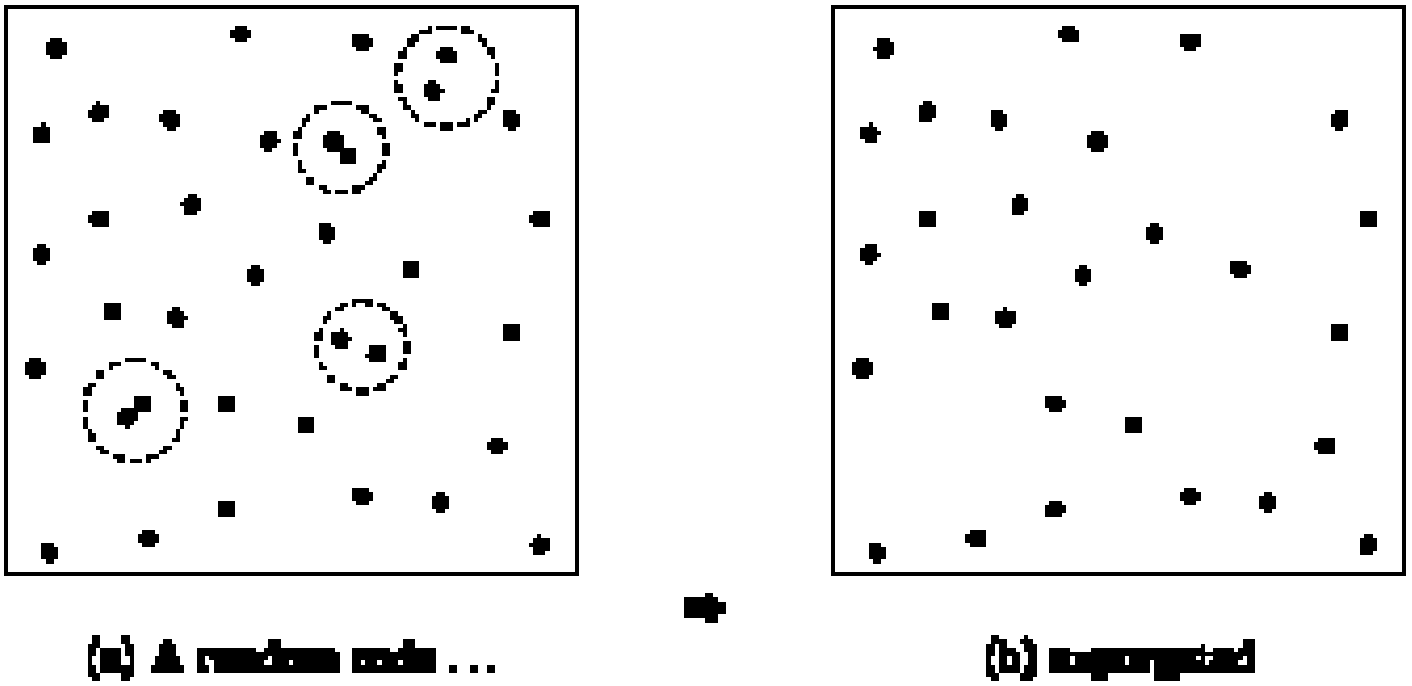
- To find the largest rate, we maximize $I(X;Y)$

$$R' < \max_{P_X} I(X;Y) - 3\delta = C - 3\delta$$

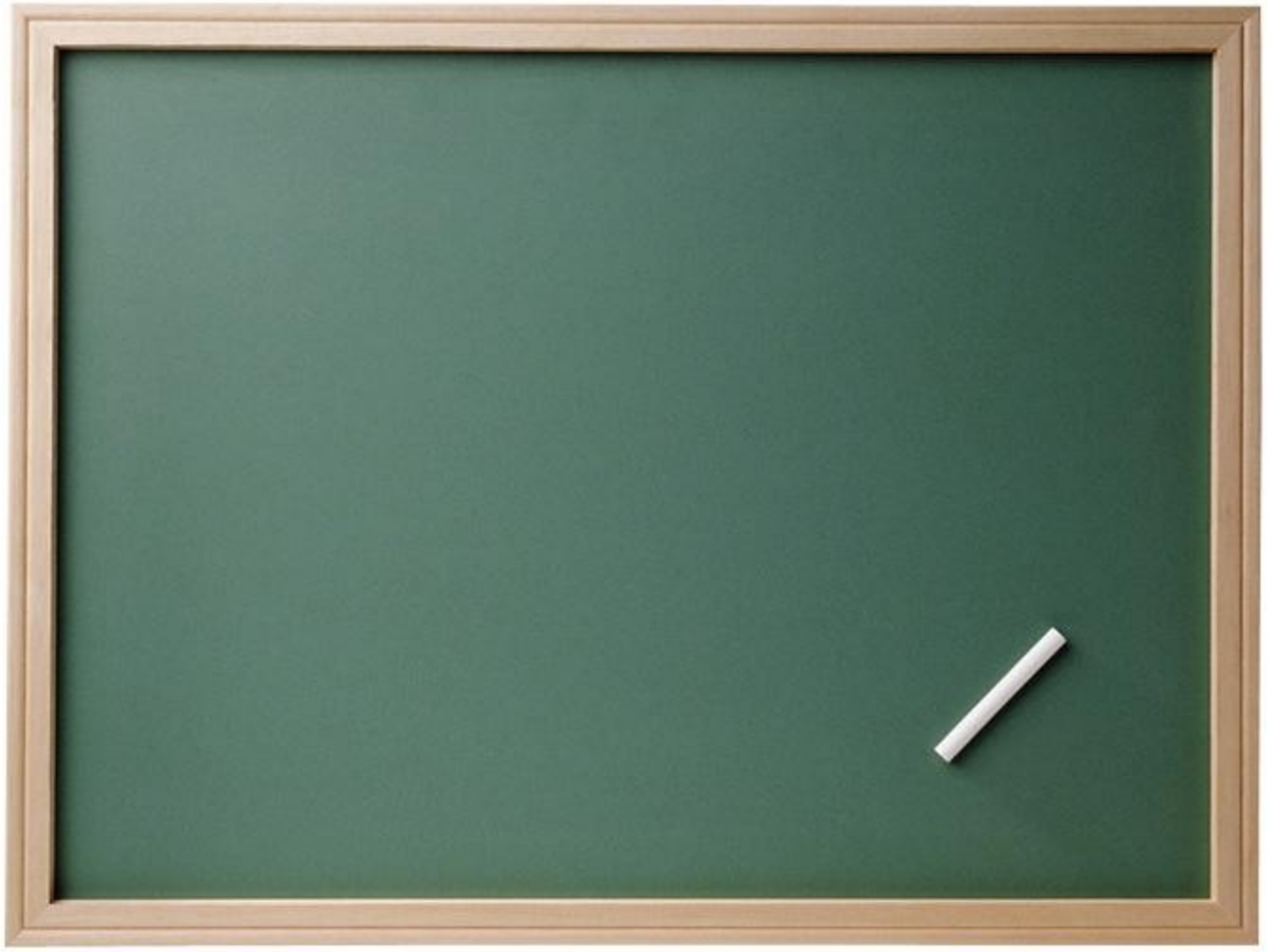
- Since the average over codes is small, there exists a code with small average block error

Average error vs Maximum error

- To find a bound on maximum error we discard some codewords which are easily confused



Bounding the maximum error



Shannon's noisy channel coding theorem (Part 1: Achievability)

- Associated with every discrete memoryless channel, there is a non-negative number C (the capacity) such that
  - For any error $\epsilon > 0$ and $R \leq C$ for large enough N, there exists a block code of length N and rate R, an a decoding algorithm, such that the maximum probability of block error is $< \epsilon$

Converse to shannon's coding theorem

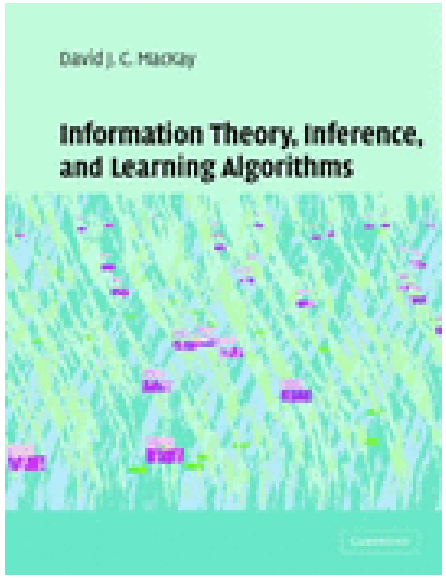
- For rates above the capacity, the error cannot be made small
- Proof during tutorial session using Fano's inequality
- Capacity forms sharp threshold for information transmission!

Next time

- Back to error correcting codes: what did we learn from Shannon's theory?

Reading for this lecture

- Chapters 9.6 to 9.8, 10.1 to 10.3 in the book



Information Theory, Inference and Learning Algorithms
by David J. C. MacKay
Cambridge University Press, 2003

- Homework due by Monday 2pm two weeks from now (Hari Raya Haji next Monday, no lecture, but tutorial sessions held)