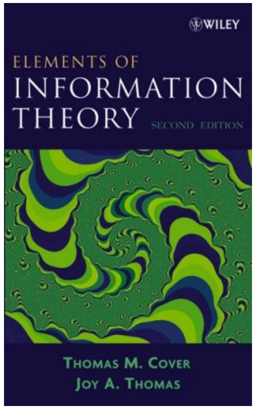
Formalization of Shannon's Theorems in SSReflect-Coq

Reynald Affeldt Manabu Hagiwara

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Motivations

incipit of:



- "Information theory answers two fundamental questions in communication theory:

  What is the ***ultimate data compression*** (answer: the entropy H), and

  what is the ***ultimate transmission rate*** of communication (answer: the channel capacity C)."

- Formalization of Shannon's theorems
  $\rightarrow$ formalization of "unconditional security"
  - One-time pad protocol
  - Key distribution protocol over a noisy channel

  $\rightarrow$ Evaluation of information leakage [Malacaria, POPL 2007; Coble, PETS 2008]

Main Contribution

- New to verification using proof-assistants:
  - Formalization of:
    - The ***source coding theorem***
      - direct part and converse part
    - The ***channel coding theorem***
      - direct part
  - ***Advanced information-theoretic notions***
    - Channels, joint typical sequences, codes, etc.

Difficulties and Approach

- Technical proofs
  - Detailed proofs appeared several years after [Shannon, 1948]
  - Quick justifications are not rare (probability theory, analytic arguments)
  - Liberal Σ/Π notations
- (Asymptotic) bounds are never made explicit
  - "this holds for n large enough"
- Plethora of concepts
  - The formalization of the many relations between information-theoretic notions is tempting…

$\Rightarrow$ Take advantage SSReflect (In particular, canonical big operators [Bertot et al., TPHOLs 2008])

$\Rightarrow$ Rework the proofs of Shannon's theorems (explicit bounds, streamlined flow)

定理 1.6 (固定長符号化逆定理) (Shannon (1948)) 分布 P を有する DMS と任意の $0 < \lambda < 1$ が与えられたとき，もし符号化率 R が

$$R < H(P)$$

を満足するならば，十分大きなブロック長 k を有する任意の固定長符号 (f, φ) は $e(f, \varphi) \geq \lambda$を満足する。 □

いま，分布 P を有する DMS と十分大きなブロック長 k の固定長符号 (f, φ) が与えられたとし，この符号について

$$R = \frac{n}{k} < H(P) \tag{1.15}$$

であるとする。この符号によって符号化したとき正しく復号される $\mathcal{X}^k$の系列の集合を $A \triangleq \{\boldsymbol{x} \in \mathcal{X}^k : \varphi(f(\boldsymbol{x})) = \boldsymbol{x}\}$ によって定めれば，式 (1.15) から明らかに

$$|A| \leq \exp(n) = \exp(kR) = \exp\{k(H(P) - \varepsilon_o)\} \tag{1.16}$$

が成り立つ。ただし，$\varepsilon_o \triangleq H(P) - R$ である。

次に，$0 < \varepsilon < \varepsilon_o$を満足する任意の$\varepsilon$が与えられたとき，集合 A に属する系列が典型系列 $B(k, \varepsilon)$ に入っているか否かで分類することで，正しく復号される確率を評価する。すなわち

$$\begin{aligned} 1 - e(f, \varphi) &= \sum_{\boldsymbol{x} \in A} P^k(\boldsymbol{x}) \\ &= \sum_{\boldsymbol{x} \in A \cap B^c(k,\varepsilon)} P^k(\boldsymbol{x}) + \sum_{\boldsymbol{x} \in A \cap B(k,\varepsilon)} P^k(\boldsymbol{x}) \\ &\leq \sum_{\boldsymbol{x} \in B^c(k,\varepsilon)} P^k(\boldsymbol{x}) + \sum_{\boldsymbol{x} \in A \cap B(k,\varepsilon)} P^k(\boldsymbol{x}) \end{aligned}$$

であり，定理 1.2 の (1) と (2) を用いれば，$k \geq k_o(P, \varepsilon)$ のとき

$$\begin{aligned} 1 - e(f, \varphi) &\leq \varepsilon + |A \cap B(k, \varepsilon)| \left\{ \max_{\boldsymbol{x} \in B(k,\varepsilon)} P^k(\boldsymbol{x}) \right\} \\ &\leq \varepsilon + |A| \exp\{-k(H(P) - \varepsilon)\} \end{aligned}$$

が得られる。この式に式 (1.16) を代入して $|A|$ を消去すると，

$$1 - e(f, \varphi) \leq \varepsilon + \exp\{-k(\varepsilon_o - \varepsilon)\}$$

となる。ここで，$\varepsilon < \varepsilon_o$に注意すれば，$k$ を十分大きくすることで $\exp\{-k(\varepsilon_o - \varepsilon)\} \leq \varepsilon$が成り立つので，結局

$$e(f, \varphi) \geq 1 - 2\varepsilon$$

が得られ，εを任意に小さく取れるので次の定理を得る。

[植松友彦,「現代シャノン理論 タイプによる情報理論」, 培風館, 1998]

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Distribution

$P_1(0)$	$P_1(1)$	$P_1(2)$	$P_1(3)$
0.3	0.2	0.3	0.2

Record dist := mkDist {
  pmf :> A $\rightarrow$ R ; *(* probability mass function*)*
  pmf0 : $\forall$a, 0 $\leq$ pmf a ;
  pmf1 : Σ_(a $\in$ A) pmf a = 1 }.

P^n : Product Distribution

P_1^2	0	1	2	3
0	0.09	0.06	0.09	0.06
1	0.06	0.04	0.06	0.04
2	0.09	0.06	0.09	0.06
3	0.06	0.04	0.06	0.04

Definition Ptuple x := Π_(i < n) P x_i

Definition Ptuple_dist
  : dist [finType of n.-tuple A].
apply mkDist with Ptuple. … Defined.

Probability of an Event

E.g., Pr P_1 [$P_1 \in$ {0,1}] = 0.5

Definition Pr P E := Σ_(a $\in$ A | E a) P a.

Typical Sequences

Intuition: Given a source of symbols, typical sequences are the most probable sequences

E.g.:

P(0)	P(1)
2/3	1/3

$\Rightarrow$ A bitstring with 2/3 of 0's would be typical

Entropy of a distribution P over A:

Definition $\mathcal{H}$:= – Σ_(a ∈ A) P a ∗ log (P a).

P, n, ε-***typical sequences*** are x tuples over A:

Definition typ_seq x :=

exp (– n ∗ ($\mathcal{H}$ P + ε)) ≤ P^n x ≤ exp (– n ∗ ($\mathcal{H}$ P – ε)).

Asymptotic Equipartition Property

- Intuition: Long enough tuples are typical

Definition aep_σ^2 := Σ_(x $\in$ A) P x $*$ (log (P x))^2 – ($\mathcal{H}$ P)^2.
Definition aep_bound ε := aep_σ^2 / P ε^3.

Lemma aep : $\forall$ ε , 0 < ε $\rightarrow$ "for n big enough" $\rightarrow$
Pr P^{n+1} [pred x | "x is not typical"] $\leq$ ε.

- Properties:

Typical sequences ($\mathcal{TS}$) are the most likely to be observed

Lemma Pr_$\mathcal{TS}$_1 : aep_bound P ε $\leq$ n+1 $\rightarrow$ Pr P^{n+1} [pred i $\in$ $\mathcal{TS}$ P n+1 ε] $\geq$ 1 - ε.

| $\mathcal{TS}$ | $\approx$ exp (n $*$ $\mathcal{H}$ P)

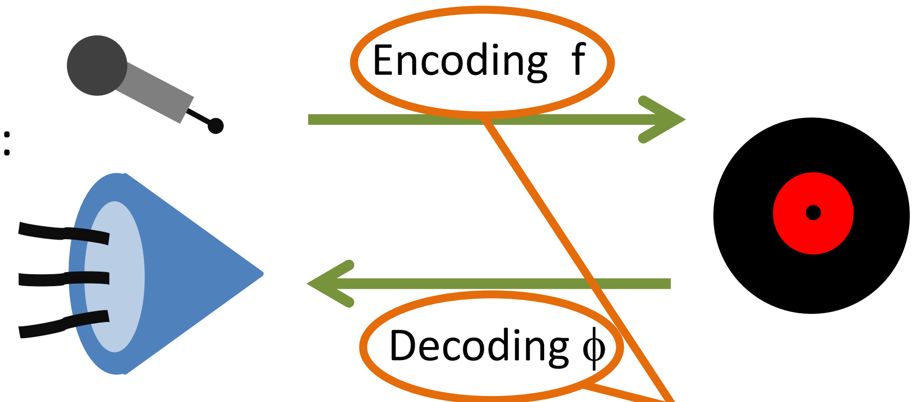
Lemma $\mathcal{TS}$_sup_inf : aep_bound P ε $\leq$ n+1 $\rightarrow$
(1 - ε) $*$ exp (n+1 $*$ ($\mathcal{H}$ P - ε)) $\leq$ | $\mathcal{TS}$ P n+1 ε | $\leq$ exp (n+1 $*$ ($\mathcal{H}$ P + ε)).

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Formalization of a Source Code

k.-tuple A symbols:
$a_1, a_2, \ldots, a_k$



n.-tuple bool:
$b_1, b_2, \ldots, b_n$

Source code:

Record scode := mkScode { enc : encT ; dec : decT }.

Source rate (objective: minimize):

Definition SrcRate (sc : scode) := n / k.

Probability of decoding error:

Definition e_{src} := Pr P^k [pred x | dec sc (enc sc x) $\neq$ x].

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Source Coding Theorem - Direct Part

"For any rate $r > \mathcal{H}P$, there is a source code with negligible error":

Theorem source_coding_direct : $\forall \lambda, 0 < \lambda < 1 \to$
$\forall r : \mathcal{Q}^+, \mathcal{H}P < r \leq 1 \to$
$\boxed{\exists k}, \boxed{\exists n}, \boxed{\exists sc : \text{scode } A\ k\ n}, r = \text{SrcRate } sc \wedge$
$e_{src}(P, sc) \leq \lambda.$

" k must be big enough"

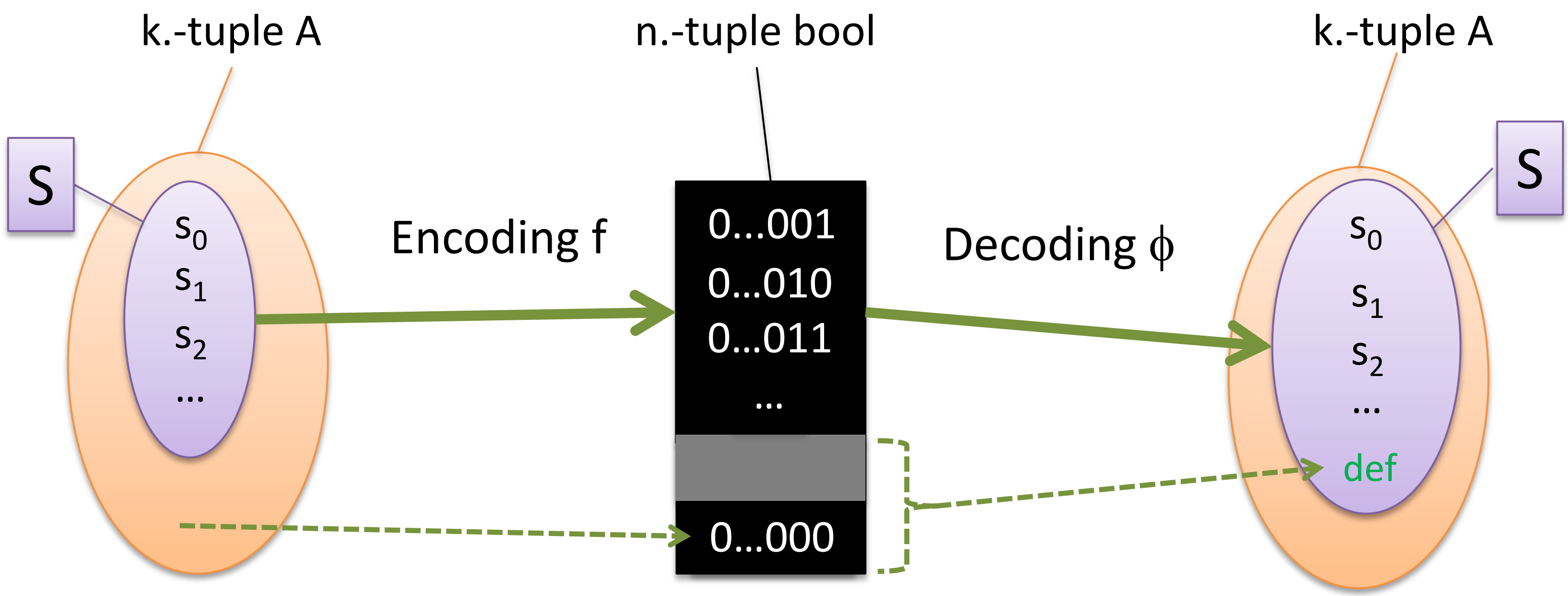
Proof sketch:

k = ? Definition ε := Rmin ($r - \mathcal{H}P$) λ.
Definition δ := Rmax (aep_bound P (ε / 2)) (2 / ε).
k must satisfy $\delta \leq k$ and $k * r$ must be a natural

n = ? n derives from k and r

sc = ? Instantiate the source code with f and ϕ from the next slide...

Source Coding Theorem – Main Idea



Source Coding Theorem – Main Idea

```
Definition f : encT A k+1 n := fun x ⇒
 if x ∈ S then
  let i := index x (enum S) in Tuple (size_nat2bin_b i+1 n)
 else
  [tuple of nseq n false].
```

```
Definition ϕ : decT A k+1 n := fun x ⇒
  let i := tuple2N x in
  if i is 0 then def else
    if i-1 < | S | then nth def (enum S) i-1 else def.
```

By construction, Lemma ϕ_f i : ϕ (f i) = i $\leftrightarrow$ i $\in$ S.

For the proof of the source coding theorem,
S := $\mathcal{TS}$, def exists because we have chosen “k big enough”

Source Coding Theorem - Converse Part

"For any rate $r < \mathcal{H}P$, all source codes have non-negligible error":

Theorem source_coding_converse : $\forall\ \lambda,\ 0 < \lambda < 1 \rightarrow$
$\forall$ r : $\mathcal{Q}^+$, $0 < r < \mathcal{H}\,P \rightarrow$
$\forall$ n k (sc : scode A k+1 n), r = SrcRate sc $\rightarrow$
$\boxed{\text{SrcConverseBound P (num r) (den r) n } \lambda \leq \text{k+1} \rightarrow}$
$e_{src}(P, sc) \geq \lambda$.

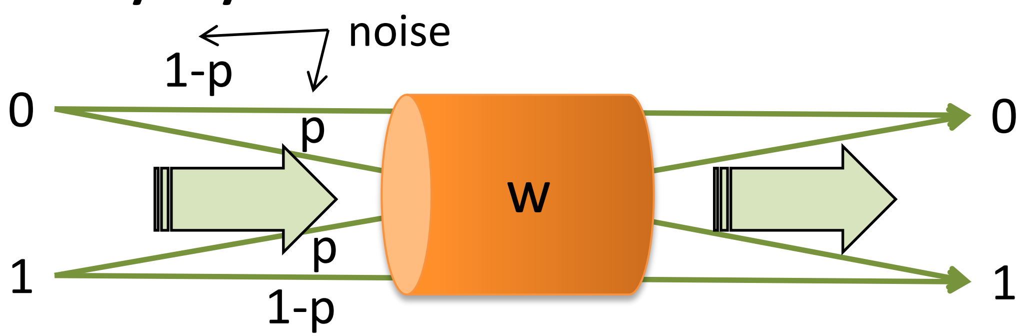
Definition ε := Rmin ((1 - λ) / 2) (($\mathcal{H}$P - r) / 2).
Definition δ := Rmin (($\mathcal{H}$P - r) / 2) (ε / 2).
Definition SrcConverseBound := Rmax (n / r)
(Rmax (aep_bound P δ) (-((log δ) / ($\mathcal{H}$P – r - δ)))).

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Formalization of a Channel

- E.g., binary symmetric channel:



- General case, channel = ***probability transition matrix***:

$$\begin{bmatrix} y_1|x_1 & y_2|x_1 & \cdots & y_{|Y|}|x_1 \\ y_1|x_2 & y_2|x_2 & \cdots & y_{|Y|}|x_2 \\ \vdots & \vdots & & \vdots \\ y_1|x_{|X|} & y_1|x_{|X|} & \cdots & y_{|Y|}|x_{|X|} \end{bmatrix}$$

Definition channel := X $\rightarrow$ dist Y.

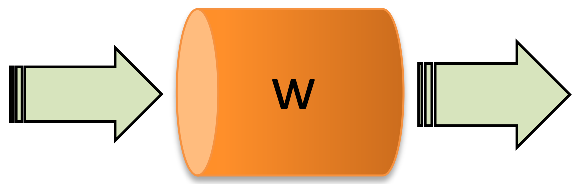
Channel Capacity

Input

Distribution P

Entropy $\mathcal{H}$ P

x_1
x_2
…
$x_{|X|}$



y_1
y_2
…
$y_{|Y|}$

Output

Distribution d(P, w)

Σ_(x $\in$ X) w x y $*$ P x

Entropy $\mathcal{H}$(P , w)

Joint

Distribution d(P ; w)

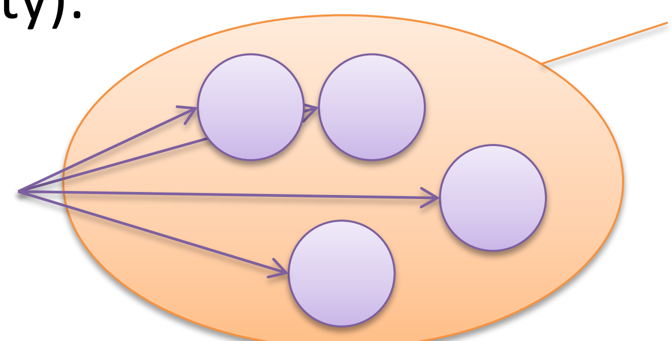
w xy.1 xy.2 $*$ P xy.1

Entropy $\mathcal{H}$(P ; w)

Definition (channel capacity):

Typical outputs
$2^{n * \mathcal{H}(P , w)}$

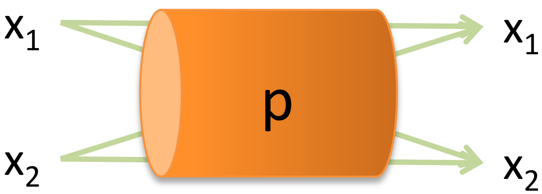
Typical outputs for one input
$2^{n * (\mathcal{H}(P ; w) - \mathcal{H} P)}$



Mutual information: $\mathcal{I}$(P ; w) = $\mathcal{H}$(P, w) – $\mathcal{H}$(P ; w) + $\mathcal{H}$ P

Definition capacity w := lub (fun P $\Rightarrow$ $\mathcal{I}$(P ; w)).

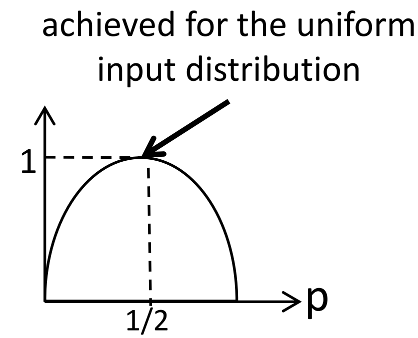
Illustration: the Binary Symmetric Channel



Binary entropy function

Definition $\mathcal{H}_2$ p :=
- p * log p – (1 - p) * log (1 - p).

Lemma $\mathcal{H}_2$_max :
$\forall$ p, 0 < p < 1 $\rightarrow$ $\mathcal{H}_2$ p $\leq$ 1



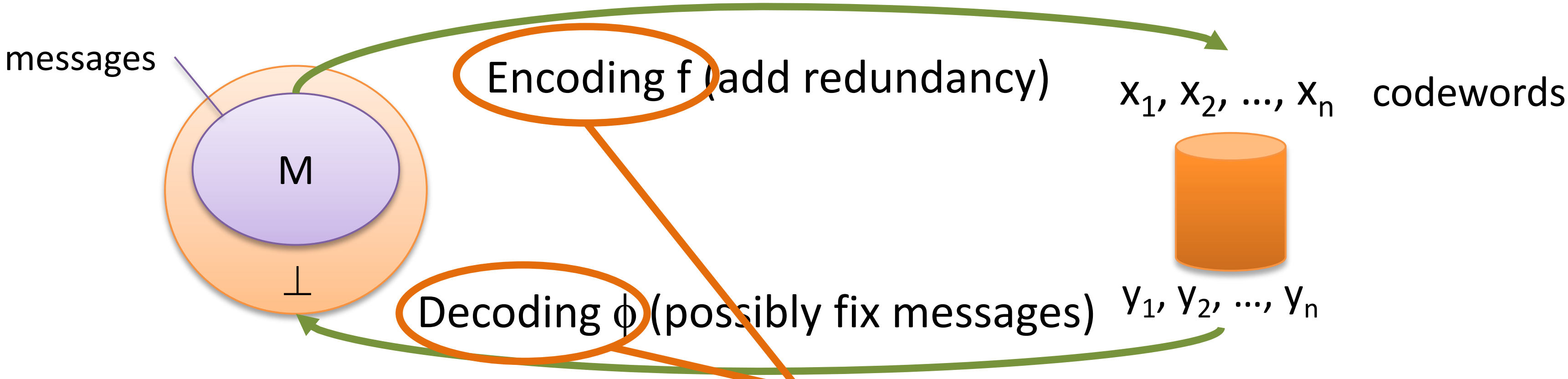
Lemma IPW : $\mathcal{I}$(P ; BSC p) = $\mathcal{H}$(P , BSC p) – $\mathcal{H}_2$ p.

Theorem BSC_capacity : capacity (BSC p) (1 – $\mathcal{H}_2$ p).

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Formalization of a Channel Code



Channel code:

Record code := mkCode { enc : encT ; dec : decT }

Rate (objective: maximize):

Definition CodeRate c := log |M| / n

Average probability of error:

Definition e_{cha} := 1/|M| ∗ Σ_(m ∈ M) e w c m.

Conditional probability of error:

Definition e w c m := Pr (w (| enc c m)) [pred y | dec x y ≠ Some m]

Outline

- Basics: Notations and typical sequences
- Source coding theorem
  - Source code formalization
  - Direct and converse proofs
- Formalization of channel capacity
- Channel coding theorem
  - Code formalization
  - Direct proof by random coding and joint typicality

Channel Coding Theorem

"For any rate r < capacity, there is a code with negligible error":

Theorem channel_coding r : CodeRateType r $\rightarrow$ r < cap $\rightarrow$
$\forall \varepsilon$, 0 < $\varepsilon \rightarrow$
$\exists$ n, $\exists$ M, $\exists$ c : code X Y M n, r = CodeRate c $\wedge$
e_{cha}(w , c) < ε.

" n must be big enough"

Proof by "random coding": we fix the decoding function ϕ and investigate all the encoding functions f

Lemma random_coding (P : dist X) w ε (ϕ : encT X M n $\rightarrow$ decT Y M n) :
Σ_(f : encT X M n) (wgth P f $*$ e_{cha}(w , mkCode f (ϕ f))) < $\varepsilon \rightarrow$
$\exists$ f, e_{cha}(w , mkCode f (ϕ f)) < ε.

Joint Typical Sequences

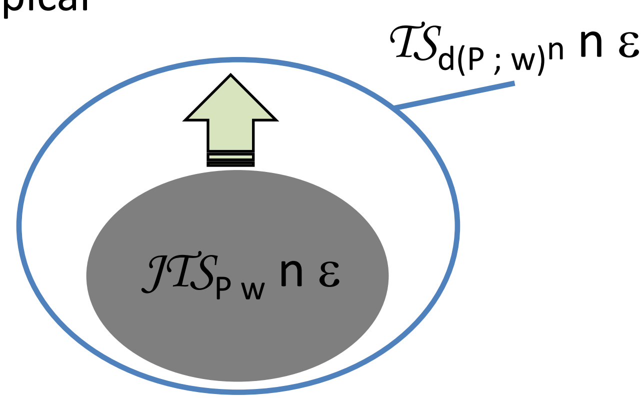
- P, w, ε, n - ***joint typical sequence***:

    Definition jtyp_seq n (xy : n.-tuple (X * Y)) ε :=

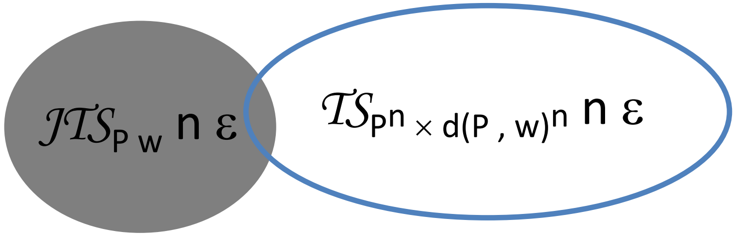
typ_seq P ε (uzip1 xy) ∧	Typical of the input
typ_seq (d(P , w)) ε (uzip2 xy) ∧	Typical of the output
typ_seq (d(P ; w)) ε xy.	Typical of the joint distribution

- Properties:

With high probability, the input and the observed output are jointly typical



With high probability, unrelated input/output are not jointly typical



Decoding by Joint Typicality

Definition jtdec P w ε (f : encT X M n) : decT Y M n :=

[ffun y ⇒ [pick m|	Pick up a message m…
((f m, y) $\in \mathcal{JTS}$ P w n ε) $\wedge$	such that (f m, y) is $\mathcal{JTS}$ and…
($\forall$ m’, m’ $\neq$ m $\rightarrow$ (f m’, y) $\notin \mathcal{JTS}$ P w n ε))]]	it is the only one in $\mathcal{JTS}$

Channel Coding theorem - Proof Sketch

$$\Sigma\_(f : encT\ X\ M\ n)\ (wgth\ P\ f * \boxed{e_{cha}(w\ ,\ mkCode\ f\ (\phi\ f))}) < \varepsilon$$

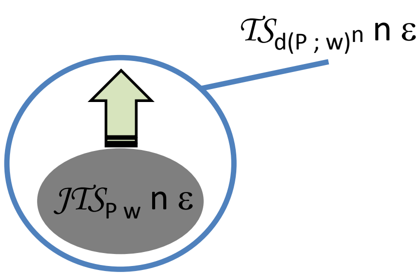
Symmetry property

$$\Pr_{(w\,(|f\,0))} \underbrace{[pred\ y \in not\_preimg\ (Jtdec\ f)\ 0]}_{\text{output y does not decode to 0}}$$

(1) the input (f 0) and the output y are not joint typical

$$\Pr_{(w\,(|f\,0))} [pred\ y \mid (f\ 0,\ y) \notin \mathcal{JTS}\ P\ w\ n\ \varepsilon\ ]$$

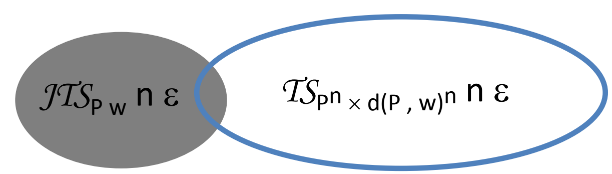
$$\Pr\ (d(P\ ;\ w)^n)\ [pred\ y \notin \mathcal{JTS}\ P\ w\ n\ \varepsilon_0\ ]$$



(2) there is another input (f i), i ≠ 0 that is joint typical with y

$$\Pr_{(w\,(|f\,0))} [pred\ y \mid (f\ i,\ y) \in \mathcal{JTS}\ P\ w\ n\ \varepsilon\ ]$$

$$\Pr\ (P^n \times d(P,w)^n)\ [pred\ y \in \mathcal{JTS}\ P\ w\ n\ \varepsilon_0\ ]$$



Conclusion

- Summary:
  - Formalization of the source coding theorem and of the (direct part of the) channel coding theorem
- Recent work (with Jonas Senizergues, ENS Cachan/AIST):
  - Converse of the channel coding theorem (with Pinsker's inequality admitted)
- Related work:
  - Mostly in HOL, based on [Hurd, PhD, 2002]
  - Probability theory
    - Expectation properties [Hasan et al., JAR 2008]
    - Weak law of large numbers [Hasan et al., ITP 2010]
  - Information theory
    - Formalization of the AEP [Mhamdi et al., ITP 2011]
    - Formalization of information leakage [Coble, PETS 2008]
- Current work:
  - Coding theory (Hamming, linear, cyclic codes $\rightarrow$ Reed-Solomon, LDPC codes?)