Multiuser information theory

An introduction

Paulo J S G Ferreira

SPL — Signal Processing Lab / IEETA
Univ. Aveiro, Portugal

April–May 2010

Contents

Contents

Contents

Contents

Multiuser information theory

- Point-to-point communication is well understood
- One source, one channel, one receiver
- Capacity regions well-known
- Rate-distortion functions well-known

Multiuser information theory

- The presence of multiple users brings new elements to the communication problem
- Instead of two main users, one has many...
- ...possibly a network of interconnected nodes — a graph
- The single-source / single-receiver case is known since Shannon
- The multiuser case includes several important problems
- Some are well understood, some are not

Multiple-access channel

- There are several senders who wish to communicate with one receiver
- The channel is shared among all senders
- This problem is well understood

Broadcast channel

- Consider now a situation in which one sender transmits to many receivers
- In this case there are problems that still lack complete answers

Relay channel

- Has one source and one receiver...
- ...but also at least one intermediate receiver/sender relay nodes
- In this case there are also problems that lack complete answers

Contents

Typical sequences

- $C = \{X_1, X_2, \ldots X_k\}$ are discrete r.v. with $p(X_1, X_2, \ldots X_k)$
- S is an ordered subset of C
- Consider n independent copies of S
- Example: n independent copies of $S = \{X_a, X_b, X_c\}$ would form a $n \times 3$ matrix:

$$\begin{bmatrix} x_{1a} & x_{1b} & x_{1c} \\ x_{2a} & x_{2b} & x_{2c} \\ x_{3a} & x_{3b} & x_{3c} \\ \vdots & \vdots & \vdots \\ x_{na} & x_{nb} & x_{nc} \end{bmatrix}$$

- The probability of a given (x_a, x_b, x_c) is

$$\prod_{i=1}^{n} p(x_{ia}, x_{ib}, x_{ic})$$

(n independent copies)

Typical sequences

- Consider n independent copies of subsets of $C = \{X_1, X_2, \dots X_k\}$
- The typical sequences $(x_1, \dots x_k)$ satisfy

$$\left| -\frac{\log p(s)}{n} - H(S) \right| < \epsilon$$

  for every $S \subset C$
- There are 2^k possible subsets of C
- The set of typical sequences is denoted by T

  (I should really use something like $T(\epsilon)$ or even $T(n, \epsilon)$, but let's keep it simple)

Law of large numbers

- We are considering k random variables
- Fix any subset A of the k variables
- Consider n independent copies A_i of it
- Then, by the law of large numbers,

$$-\frac{\log p(A_1, A_2, \ldots A_n)}{n} = -\frac{1}{n}\sum \log p(A_i) \to H(A)$$

- Compare with the definition of typical sequence
- The law of large numbers guarantees that as $n \to \infty$ the conditions are satisfied, no matter how small ϵ is

Notation

- $T(S)$ is the restriction of T to a subset S of the k vectors
- Let for example $S = (X_a, X_b)$
- In this case, the typicality condition would mean

$$\left| -\frac{\log p(x_a, x_b)}{n} - H(x_a, x_b) \right| < \epsilon$$

$$\left| -\frac{\log p(x_a)}{n} - H(x_a) \right| < \epsilon$$

$$\left| -\frac{\log p(x_b)}{n} - H(x_b) \right| < \epsilon$$

More notation

- Consider the condition

$$\left| -\frac{\log p(a)}{n} - H(A) \right| < \epsilon$$

- Rewrite it as

$$-\frac{\log p(a)}{n} - H(A) = \theta\epsilon, \qquad |\theta| \leq 1$$

- Solve for $p(a)$:

$$p(a) = 2^{-n(H(A)+\theta\epsilon)}$$

- This is

$$2^{-n(H(A)+\epsilon)} \leq p(a) \leq 2^{-n(H(A)-\epsilon)}$$

- This range is abbreviated as follows:

$$\boxed{p(a) = 2^{-n(H(A)\pm\epsilon)}}$$

Some basic results (i)

- If n is sufficiently large, then for any subset S of $C = \{X_1, X_2, \dots X_k\}$

$$\boxed{P(T(S)) \geq 1 - \epsilon}$$

- This is a direct consequence of the law of large numbers

Some basic results (ii)

- The probability of any sequence s in $T(S)$ is

$$\boxed{p(s) = 2^{-n(H(S)\pm\epsilon)}}$$

- This is a consequence of the definition

$$\left| -\frac{\log p(s)}{n} - H(S) \right| < \epsilon$$

and the notation introduced

Some basic results (iii)

- The size of $T(S)$ can be estimated by

$$\boxed{|T(S)| = 2^{n(H(S)\pm 2\epsilon)}}$$

- First, use the previous result as follows:

$$\begin{aligned} 1 &\geq \sum_{s \in T(S)} p(s) \\ &\geq \sum_{s \in T(S)} 2^{-n(H(S)+\epsilon)} \\ &\geq |T(S)| 2^{-n(H(S)+\epsilon)} \end{aligned}$$

- Thus, $|T(S)| \leq 2^{n(H(S)+\epsilon)}$

Some basic results (iii)

- Also need an inequality in the opposite direction:

$$
\begin{aligned}
1-\epsilon &\le \sum_{s \in T(S)} p(s) \\
&\le \sum_{s \in T(S)} 2^{-n(H(S)-\epsilon)} \\
&\le |T(S)| 2^{-n(H(S)-\epsilon)}
\end{aligned}
$$

- This shows that

$$|T(S)| \ge (1-\epsilon) 2^{n(H(S)-\epsilon)} \ge 2^{-n\epsilon} 2^{n(H(S)-\epsilon)}$$

- Hence, $2^{n(H(S)-2\epsilon)} \le |T(S)| \le 2^{n(H(S)+\epsilon)}$, as wanted

Some basic results (iv)

- The conditional probability of a, b in $T(A, B)$ can also be estimated:

$$\boxed{p(a|b) = 2^{-n(H(A|B)\pm 2\epsilon)}}$$

- Proof: by the previous results

$$p(b) = 2^{-n(H(B)\pm\epsilon)} \qquad p(a, b) = 2^{-n(H(A,B)\pm\epsilon)}$$

- The result follows from

$$p(a|b) = \frac{p(a, b)}{p(b)} = 2^{-n(H(A,B)\pm\epsilon)} 2^{n(H(B)\pm\epsilon)}$$

and

$$H(A, B) = H(B) + H(A|B)$$

Consequences

- Let S_1, S_2 be subsets of $C = \{X_1, X_2, \dots X_k\}$
- Similar idea shows that the number of sequences that are jointly typical with a given (typical) sequence $g \in S_2$ satisfies
$$|T(S_1|g)| \leq 2^{n(H(S_1|S_2)+2\epsilon)}$$
- For three subsets of C, the probability of joint typicality is

$$\boxed{P \equiv P[(S_1, S_2, S_3) \in T] = 2^{-n(I(S_1;S_2|S_3)\pm 6\epsilon)}}$$

Why?

- S_1, S_2, S_3 are independent given S_3, same pairwise marginals
- Thus
  $p(S_1, S_2, S_3) = p(S_1, S_2|S_3)p(S_3) = p(S_1|S_3)p(S_2|S_3)p(S_3)$

$$
\begin{aligned}
P &= \sum p(s_3)p(s_1|s_3)p(s_2|s_3) \\
&= |T(S_1, S_2, S_3)| 2^{-n(H(S_3)\pm\epsilon)} 2^{-n(H(S_1|S_3)\pm 2\epsilon)} 2^{-n(H(S_2|S_3)\pm 2\epsilon)} \\
&= 2^{n(H(S_1,S_2,S_3)\pm\epsilon)} 2^{-n(H(S_3)\pm\epsilon)} 2^{-n(H(S_1|S_3)\pm 2\epsilon)} 2^{-n(H(S_2|S_3)\pm 2\epsilon)}
\end{aligned}
$$

- Result follows from
  $I(X, Y|Z) = -H(X, Y, Z) + H(Z) + H(X|Z) + H(Y|Z)$

Why?

$$\begin{aligned}
I(X,Y|Z] &= \sum p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}\\
&= \sum p(x,y,z)\log\frac{p(x,y,z)}{p(z)p(x|z)p(y|z)}\\
&= -H(X,Y,Z) + H(Z) + H(X|Z) + H(Y|Z)
\end{aligned}$$

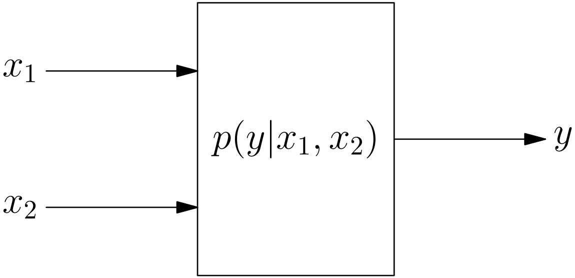
Contents

Multiple access channel

- Simplest version: two senders, one receiver
- New difficulty: in addition to noise there is interference
- Arises in several practical problems
  - One satellite receiver, many emitting ground stations
  - One base station, many cell phones within its range

Multiple access channel

- Characterised by three alphabets
- One set of probabilities $p(y|x_1, x_2)$



Code

- Codes for this channel must account for x_1 and x_2
- Notation: $\left(2^{nR_1}, 2^{nR_2}, n\right)$
- Need:
  - Two encoding functions
  - One decoding function $g(\cdot)$
- The functions deal with integers in the range
  - $\{1, \ldots, 2^{nR_1}\}$, for x_1
  - $\{1, \ldots, 2^{nR_2}\}$, for x_2

Operation

- Each sender picks an index w_1, w_2 in the appropriate range and sends the corresponding codeword
- Probability of error:

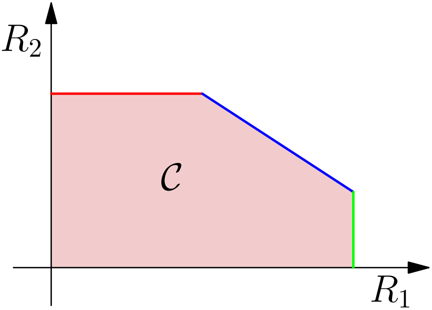
$$P_e^n = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1,w_2} P\left[g(Y^n) \neq (w_1,w_2) | (w_1,w_2)\text{sent}\right]$$

- A pair (R_1, R_2) is achievable if there exists a sequence of codes $\left(2^{nR_1}, 2^{nR_2}, n\right)$ such that $P_e^n \to 0$
- The capacity region is the closure of the set of achievable pairs

Capacity region

The capacity region is described by

$$R_1 < I(X_1, Y|X_2)$$
$$R_2 < I(X_2, Y|X_1)$$
$$R_1 + R_2 < I(X_1, X_2|Y)$$



Independent binary symmetric channels

- In this case one can send at rate
  - $1 - H(p_1)$ over the first channel
  - $1 - H(p_2)$ over the second channel
- There is no interference
- The capacity region is bounded by the vertical and horizontal lines only

Binary multiplier channel

- In this case $Y = X_1 X_2$
- By setting $X_1 = 1$ one can achieve $R_2 = 1$ on the second channel
- By setting $X_2 = 1$ one can achieve $R_1 = 1$ on the first channel
- It is impossible to make $R_1 + R_2 > 1$ since Y is binary
- Time-sharing makes $R_1 + R_2 = 1$ possible
- The capacity region is bounded by the $R_1 + R_2 = 1$ diagonal line

Erasures and multiple access

- Consider $Y = X_1 + X_2$, binary inputs, ternary output
- Value $Y = 1$ can arise as a result of $(0, 1)$ or $(1, 0)$
- By setting $X_1 = 0$ one can achieve $R_2 = 1$ on the second channel
- By setting $X_2 = 0$ one can achieve $R_1 = 1$ on the second channel
- If $R_1 = 1$, the channel looks like an erasure channel to the other channel
- The capacity of the erasure channel with $p = 1/2$ is $1/2$
- Hence if $R_1 = 1$ one can send an additional $1/2$ bit of information on the other channel

Achievability proof

- To construct the codebook, generate:
  - 2^{nR_1} codewords $X_1(i), i = 1, 2, \ldots 2^{nR_1}$
  - 2^{nR_2} codewords $X_2(i), i = 1, 2, \ldots 2^{nR_2}$
- Each element is generated independently following $p_1(x_1)$ ($p_2(x_2)$)
- To send i, sender 1 sends $X_1(i)$
- To send j, sender 2 sends $X_2(j)$
- The decoder picks (i, j) such that $(x_1(i), x_2(j), y)$ are typical, if this is possible (that is, if the pair exists and is unique)
- Otherwise there is an error
- By symmetry, it is possible to assume w.l.g. that $(i, j) = (1, 1)$

Proof outline

- Let T be the set of typical (x_1, x_2, y) sequences
- There is an error if:

  (a) The correct codewords are not typical with the received sequence

  or if

  (b) There are incorrect codewords typical with the received sequence

- The error probability is the sum of the probabilities of (a) and (b)

Error probability

- The probabilities can be estimated using the previous results on typicality
- The probability of (a) is the probability of $(x_1(1), x_2(1), y) \notin T$
- Roughly speaking, this goes to zero as $n \to \infty$ by the law of large numbers and the definition of typicality
- See the “basic results” section

Error probability

- The probability of (b) is the probability of existence of a pair (i, j) other than $(1, 1)$ typical with y
- This is bounded by the sum of the probabilities of $(x_1(i), x_2(j), y)$ with $(i, j) \neq (1, 1)$ being typical
- Roughly speaking, this can be bounded by terms similar to $2^{-nI(X_1, X_2|Y)}$
- These terms, as seen before, also tend to zero as n increases
- See the “consequences” sections

Convexity

- The capacity region $\mathcal{C}$ is convex:
- Let $(a, b) \in \mathcal{C}$
- Let $(c, d) \in \mathcal{C}$
- Let $\lambda \in [0, 1]$
- Then $z(\lambda) = \lambda(a, b) + (1 - \lambda)(c, d)$ is also in $\mathcal{C}$
- This means that
  - Source 1 may send at the rate $\lambda a + (1 - \lambda)c$, if it may send at rates a and c
  - Source 2 may send at the rate $\lambda b + (1 - \lambda)d$, if it may send at rates b and d
- Of course, geometrically $z(\lambda)$ is a point in the line segment joining (a, b) to (c, d)

Convexity proof

- Given: two codebooks at the rates (a, b) and (c, d)
- Construct: a code at the rate

$$(\lambda a + (1-\lambda)c, \lambda b + (1-\lambda)d)$$

- How?
  - Use codebook 1 for the first λn symbols
  - Use codebook 2 for the remaining $(1-\lambda)n$ symbols
- Note each code being used a fraction λ or $1-\lambda$ of the time
- The new code has rates $\lambda(a, b) + (1-\lambda)(c, d)$
- The error probability for the new code is bounded by the sum of the two error probabilities for the two segments
- Hence, it also goes to zero: the new point is in the capacity region