noisy channels

discussion

This is a summary of chapter 5 in "Codes" text, using their notation.

The math is conditional probability and entropy again, but this time with a different interpretation.

We're back to thinking about 1st order Markov again, for now, but with an added "noise" source between the sender and receiver.

symbol ***p*** sent → NOISE → symbol ***q*** received

p ≡ $[p_1, p_2, \ldots]$ = probabilities of symbols produced by sender = (1 x N) row vector

q ≡ $[q_1, q_2, \ldots]$ = probabilities symbols seen by receiver = (1 x N) row vector

$\mathbf{\Gamma} \equiv \left[\Gamma_{ij}\right]$ = [conditional probablity(got j | sent i)]

= probability of receiving symbol j given that i was sent = (N x N matrix)

= prob(**q** | **p**)

where i = row = sent symbol, and j = column = received symbol

$\mathbf{t} \equiv \left[t_{ij}\right] = \Gamma_{ij}\, p_i$ =prob(get j| sent i) prob(sent i) = prob(get j & sent i) = (N x N matrix)

= probability(**p** & **q**) = [probability(sent i, got j)] = probability(i,j)

= joint probability distribution

We can also define the conditional probability in the other direction

$\mathbf{\Delta} \equiv \left[\Delta_{ij}\right]$ = [conditional probability(sent i | got j)]

$= \left[t_{ij} / q_j \right]$ = N x N matrix

For any probability distribution $\boldsymbol{g}$ that sums to 1, we already have a notion of entropy :

$$H(\boldsymbol{g}) \equiv -\sum_i g_i \log_2(g_i)$$

We now define a conditional entropy.

The text uses two similar notations, and shows that two definitions are equivalent.

$$H(\boldsymbol{p} \mid \boldsymbol{q}) \equiv H(\boldsymbol{p} \,\&\, \boldsymbol{q}) - H(\boldsymbol{q}) \equiv H(\mathbf{\Gamma};\ \boldsymbol{p})$$

$$= H(\boldsymbol{t}) - H(\boldsymbol{q})$$

$$= \sum_j q_j\, H(\boldsymbol{p} \mid j) = -\sum_j q_j \sum_i \left(t_{ij}/q_j\right) \log_2\left(t_{ij}/q_j\right)$$

$$= -\sum_{ij} t_{ij} \log_2\left(t_{ij}/q_j\right)$$

Likewise,

$$H(\boldsymbol{q} \mid \boldsymbol{p}) \equiv H(\boldsymbol{p} \;\&\; \boldsymbol{q}) - H(\boldsymbol{p}) \equiv H(\boldsymbol{\Delta}\,;\, \boldsymbol{q})$$
$$= H(\boldsymbol{t}) - H(\boldsymbol{p})$$
$$= \sum_i p_i\, H(\boldsymbol{q} \mid i) = -\sum_i p_i \sum_j \left(t_{ij} / p_i\right) \log_2\left(t_{ij} / p_i\right)$$
$$= -\sum_{ij} t_{ij} \log_2\left(t_{ij} / p_i\right)$$

The "channel capacity" of the source **p** through the channel **Γ** is then defined as the difference of two of these entropies :

$$f_\Gamma(\boldsymbol{p}) \equiv H(\boldsymbol{p}) - H(\Gamma;\, \boldsymbol{p}) = H(\boldsymbol{p}) - H(\boldsymbol{p} \mid \boldsymbol{q})$$

= bits of information about **p** available to someone who knows **q** .

All this implies (from the definitions of conditional probabilities) a bunch of identities :

$1 = \sum_j \Gamma_{ij}$	for every i	sum of each conditional probability is 1
$1 = \sum_i \Delta_{ij}$	for every j	ditto
$1 = \sum_{ij} t_{ij}$		sum of all joint probabilities is 1
$q_j = \sum_i p_i \Gamma_{ij}$	or	$\mathbf{q} = \mathbf{p}\,\mathbf{\Gamma}$ using matrix multiplication
$p_i = \sum_j t_{ij}$		p marginal probability from t
$q_j = \sum_i t_{ij}$		q marginal probability from t

Example :

$\mathbf{p} = [0.7, 0.3]$ = [prob(send 0), prob(sending 1)]

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix} = \begin{bmatrix} \text{prob(get 0} \mid \text{send 0)} & \text{prob(get 1} \mid \text{send 0)} \\ \text{prob(get 0} \mid \text{send 1)} & \text{prob(get 1} \mid \text{send 1)} \end{bmatrix} = \begin{bmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{bmatrix}$$

$$\mathbf{q} = [0.7, 0.3] \begin{bmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{bmatrix} = [0.699, 0.301] = [\text{prob(get 0)}, \text{prob(get 1)}]$$

$$\mathbf{t} = \begin{bmatrix} p_0\,\Gamma_{00} & p_0\,\Gamma_{01} \\ p_1\,\Gamma_{10} & p_1\,\Gamma_{11} \end{bmatrix} = \begin{bmatrix} 0.693 & 0.007 \\ 0.006 & 0.294 \end{bmatrix}$$

Everything else is calculated below, including the various entropies.

■ definitions

```
H[prob_] := Total[-Flatten[prob] Log[2, Flatten[prob]]];
H2[pXY_] := Block[{},
   pX = Total[Transpose[pXY]];
   Return[Sum[-pXY[[i, j]] Log[2, pXY[[i, j]] / pX[[j]]]]];
  ];
```

```
Γ = {{0.99, 0.01}, {0.02, 0.98}};
p = {0.7, 0.3};
q = p . Γ;
t = Table[p[[i]] Γ[[i, j]], {i, 1, 2}, {j, 1, 2}];
Δ = Table[t[[i, j]] / q[[j]], {i, 1, 2}, {j, 1, 2}];
```

■ check

```
Γ // MatrixForm
```

$$\begin{pmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{pmatrix}$$

```
p // MatrixForm
```

$$\begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix}$$

```
q // MatrixForm
```

$$\begin{pmatrix} 0.699 \\ 0.301 \end{pmatrix}$$

```
t // MatrixForm
```

$$\begin{pmatrix} 0.693 & 0.007 \\ 0.006 & 0.294 \end{pmatrix}$$

```
Δ // MatrixForm
```

$$\begin{pmatrix} 0.991416 & 0.0232558 \\ 0.00858369 & 0.976744 \end{pmatrix}$$

```
Total[Δ] // MatrixForm
```

$$\begin{pmatrix} 1. \\ 1. \end{pmatrix}$$

```
Total[p]
```

1.

```
Total[q]
```

1.

```
Total[Transpose[Γ]] // MatrixForm
```

$$\begin{pmatrix} 1. \\ 1. \end{pmatrix}$$

```
Total[Total[t]]
```

```
1.
```

```
Total[t]
```

```
{0.699, 0.301}
```

```
Total[Transpose[t]]
```

```
{0.7, 0.3}
```

■ entropies

Entropy of input

```
H[p]
```

```
0.881291
```

Entropy of output

```
H[q]
```

```
0.88251
```

Entropy of joint distribution, knowing everything

```
H[t]
```

```
0.980278
```

H (**q** | **p**) calculated two ways :

1. Prob(got| sent 0)

```
Γ[[1]]
```

```
{0.99, 0.01}
```

Prob(got| sent 1)

```
Γ[[2]]
```

```
{0.02, 0.98}
```

Entropy of those two conditionals

```
H[Γ[[1]]]
```

```
0.0807931
```

```
H[Γ[[2]]]
```

```
0.141441
```

weighted average of the conditional entropies

```
p . {H[Γ[[1]]], H[Γ[[2]]]}
```

```
0.0989874
```

2. difference of (joint entropy - source entropy)

```
H[t] - H[p]
```

```
0.0989874
```

H(**p** | **q**) = H(Γ; **p)** calculated two ways :

```
H[t] - H[q]
```

```
0.0977684
```

```
q . {H[Transpose[Δ][[1]]], H[Transpose[Δ][[2]]]}
```

```
0.0977684
```

■ channel capacity

```
H[p] - (H[t] - H[q])
```

```
0.783523
```

The interpretation for these numbers then is :
a) The source **p** with (0,1) probabilities of (0.7, 0.3), without noise, can deliver 0.88 bits of information per bit sent.
b) With this noisy Γ channel, which flips (0,1) with (0.01, 0.02) probability, only 0.78 bits of information per bit sent is possible.