Q1 Requiring one bit above entropy

Monday, August 29, 2011 10:20 AM

We want to come up with some simple example. Therefore we shall start by considering the Bernoulli RV X which has only two possible values:

Consider $X \sim \text{Bernoulli}(p_1)$: $p(x) = \begin{cases} p_1, & x = 1, \\ 1-p_1, & x = 0, \\ 0, & \text{otherwise.} \end{cases}$

Because there are only two possible values, the pairing in the Huffman coding process must be between these two values:

x	$p(x)$		$c(x)$	$\ell(x)$
0	$1-p_1$	0	0	1
1	p_1	1	1	1

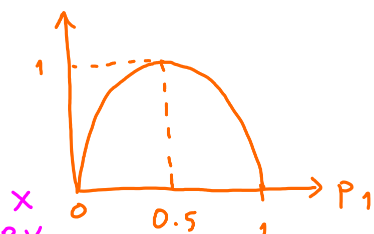
Therefore, Huffman coding (without extension) always needs 1 bit per symbol.

Now, we want the $\underbrace{\text{expected length}}_{=1}$ to be $\approx H(X)+1$

So, we must have $H(X) \approx 0$.

By definition, $H(X) = -\sum_x p(x) \log_2 p(x) = -(1-p_1)\log(1-p_1) - p_1 \log p_1$

You may recall that in Section 2.4, this expression is called the binary entropy function. The plot of this function is shown in the lecture notes; it is sketched below:



Note that to get this function to be ≈ 0, we need to consider $p_1 \approx 0$ or $p_1 \approx 1$.

Also note that we don't want to have $p_1 = 0$ or $p_1 = 1$ because they would make our RV X degenerated (deterministic). For degenerated RV, we don't have to waste any bit to convey its value. (The value is already pre-determined.) For example, if $p_1 = 1$, we know that $P[X=1] = 1$ and hence $X \equiv 1$ all the time. There is no uncertainty. Anyone can guess the value of X with 100% accuracy by simply guessing the value 1 every time. Alternatively, we may think of the situation here as sending the empty string (ε) to the receiver. So, $\mathbb{E}[\ell(X)] = \mathbb{E}[0] = 0$. Because $H(X)$ is also 0, we have $\mathbb{E}[\ell(X)] = H(X)$ and not $H(X)+1$.

More formally we can take $\lim_{p_1 \to 0}$ or $\lim_{p_1 \to 1}$ on the function and show that $H(X) \to 0$

$$\left(\lim_{x \to 0^+} x \ln x = \lim_{x \to 0^+} \frac{\frac{1}{x}}{\frac{-1}{x^2}} = \lim_{x \to 0^+} (-x) = 0 \right)$$

$$\left(\lim_{x \to 0^+} x \ln x = \underset{\uparrow}{=} \lim_{x \to 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x \to 0^+} (-x) = 0 \right)$$

L'Hôpital's rule

So, if p_1 is close to 1 or 0, the entropy will be almost 0.

But we still need $\mathbb{E}[\ell(X)] = 1$ bit to send X.

So, the expected length will be very close to $H(X) + 1$.

Q2 ML and MAP detectors for repetition code

Tuesday, September 20, 2011 4:36 PM

Recall that the code rate of a repetition code is $\frac{1}{n}$ where n is the number of times that an info-bit is repeated.

Here, the code rate is $\frac{1}{5}$. So, $n = 5$.

Codebook :

s	$\underline{x}$
0	00000
1	11111

usually, for block coding, this part is also a row vector. This is because we work with k info-bit at a time. In which case, $\underline{s}$ would be a row vector of length k. However, for repetition code, we have $k=1$.

(a) In class, we have shown that for repetition code over BSC with $p<0.5$, the ML decoder is the same as a majority-vote decoder:

$$\hat{s}_{ML}(\underline{y}) = \begin{cases} 0 & \text{if } n_0 > n_1, \\ 1 & \text{if } n_0 < n_1. \end{cases}$$

n_1 = the number of 1's in the observed vector $\underline{y}$

n_0 = the number of 0's in the observed vector $\underline{y}$

Therefore,

n_0	n_1	$\hat{s}_{ML}(\underline{y})$
0	5	1
1	4	1
2	3	1
3	2	0
4	1	0
5	0	0

$$P(\mathcal{E}) = \text{the probability that the } \# \text{ bits flipped by the BSC } \underbrace{\geq 3}_{\text{more than half}}$$

$$= \sum_{k=3}^{5} (\text{the probability that the } \# \text{ bits flipped by the BSC} = k)$$

$$= \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5 (1-p)^0$$

$$= 10 p^3 (1-p)^2 + 5 p^4 (1-p) + p^5$$

$= 0.317$

$p = 0.4$

(b) In class, we have shown that the MAP detector for block coding over the BSC (with crossover probability p) is given by

$$\hat{\underline{x}}_{MAP}(\underline{y}) = \arg\max_{\underline{x}} \left(\frac{p}{1-p}\right)^{d(\underline{x},\underline{y})} p(\underline{x}).$$

We have also shown that when the block coding is done simply by using repetition code,

$$\hat{\underline{x}}_{MAP}(\underline{y}) = \begin{cases} \underline{0} & \text{when } \left(\frac{p}{1-p}\right)^{n_1-n_0} \frac{p_0}{p_1} > 1 \\ \underline{1} & \text{when } \left(\frac{p}{1-p}\right)^{n_1-n_0} \frac{p_0}{p_1} < 1 \end{cases}$$

$p = 0.4$

$\frac{p}{1-p} = \frac{0.4}{0.6} = \frac{2}{3}$ $\qquad \frac{p_0}{p_1} = \frac{0.4}{0.6} = \frac{2}{3}$

n_1	n_0	$n_1 - n_0$	$\left(\frac{p}{1-p}\right)^{n_1-n_0}$	$\times \frac{p_0}{p_1}$	$\hat{S}_{MAP}(\underline{y})$
0	5	-5	$(2/3)^{-5}$	>1	0
1	4	-3	$(2/3)^{-3}$	>1	0
2	3	-1	$(2/3)^{-1}$	$=1$	Any
3	2	1	$(2/3)$	<1	1
4	1	3	$(2/3)^3$	<1	1
5	0	5	$(2/3)^5$	<1	1

If this is set to be 0, then it is the same as $\hat{S}_{ML}$.

So, can set $\hat{S}_{MAP}(y) = \hat{S}_{ML}(y)$ above.

In which case, same probability of error.

Remark 1 Let's try to set this to be 1 so that the detector is different from $\hat{S}_{ML}$.

Then, $P[\hat{S} \neq S \mid S=0] = P[\hat{S} \neq 0 \mid S=0] = \sum_{k=2}^{5} p^k (1-p)^{5-k} = 0.663$

$P[\hat{S} \neq S \mid S=1] = P[\hat{S} \neq 1 \mid S=1] = \sum_{k=4}^{5} p^k (1-p)^{5-k} = 0.087$

So, $P[\mathcal{E}] = P[\hat{S} \neq S \mid S=0]P[S=0] + P[\hat{S} \neq S \mid S=1]P[S=1]$

$\approx 0.663 \times 0.4 + 0.087 \times 0.6 = 0.317 \rightarrow$ same performance!

Remark (2) The "any" or "don't care" or "does not matter" in the table means that the decoder can choose to guess 1 or 0 without changing the overall error-rate performance.

However, some decoder may issue a "decoding error" warning at the receiver and avoid making a guess.

(We will not discuss this type of decoder.)

(c)

(i) $P[S=0 \mid \underline{Y}=01001] = \dfrac{P[\underline{Y}=01001 \mid S=0]\, P[S=0]}{P[\underline{Y}=01001]} = \dfrac{p^2(1-p)^3 \times p_0}{P[\underline{Y}=01001]}$ $\quad$ $P[S=0]=0.45$

conditioned on $S=0$, the repetition code would be 00000 and therefore, $\underline{Y}=01001$ means the BSC flips exactly two bits.

By the total probability theorem,

$$P[\underline{Y}=01001] = P[\underline{Y}=01001 \mid S=0]\, P[S=0] + P[\underline{Y}=01001 \mid S=1]\, P[S=1]$$

$$= p^2(1-p)^3 p_0 + p^3(1-p)^2 p_1 \approx 0.0282$$

$P[S=1] = 0.55$

conditioned on $S=1$, the repetition code would be 11111 and therefore $\underline{Y}=01001$ means the BSC flips exactly 3 bits.

Therefore, $P[S=0 \mid \underline{Y}=01001] \approx 0.5510$

(ii) $P[S=1 \mid \underline{Y}=01001] = 1 - P[S=0 \mid \underline{Y}=01001] = 0.4490$

(iii) $P[S=0 \mid \underline{Y}=01001] > P[S=1 \mid \underline{Y}=01001]$.

Therefore, it is more likely that $S=0$ was transmitted.

(d) (i) $P[S=0 \mid \underline{Y}=01001] = \dfrac{p^2(1-p)^3 \times p_0}{p^2(1-p)^3 p_0 + p^3(1-p)^2 p_1} = \dfrac{1}{1+\frac{p}{1-p}\frac{p_1}{p_0}}$

$$= \frac{1}{1+\frac{2}{3}\times\frac{1-p_0}{p_0}} = \frac{3p_0}{p_0+2}$$

$1 + \frac{}{3} \times \frac{}{p_0}$

(ii) $P[S=1 \mid \underline{Y} = 01001] = 1 - \qquad = \frac{2-2p_0}{p_0+2}$

(iii) $P[S=0 \mid \underline{Y} = 01001]$ is more likely iff

$$\frac{3p_0}{p_0+2} > \frac{2-2p_0}{p_0+2}$$

$$p_0 > \frac{2}{5} = 0.4$$

$P[S=1 \mid \underline{Y} = 01001]$ is more likely iff $p_0 < 0.4$

When $p_0 = 0.4$, the conditional probabilities are the same and the two cases are equally likely.

Majority voting would always guess 0 because the #0s in 01001 is greater than the #1s. This would agree with our answer here when $p_0 > 0.4$.

Q3 ML and MAP detectors for block coding

Tuesday, October 09, 2012 9:50 AM

(a) We can find the most likely transmitted codeword by using the MAP decoder

01001

$$\hat{\underline{x}}_{MAP}(\underline{y}) = \arg\max_{\underline{x}} P[\underline{X} = \underline{x} \mid \underline{Y} = \underline{y}]$$

$$= \arg\max_{\underline{x}} Q(\underline{y} \mid \underline{x})\, p(\underline{x})$$

Because all the codewords are equally likely. The term $p(\underline{x})$ in the above expression does not depend on $\underline{x}$ and hence it can be eliminated. In other words, when the codewords are equally likely,

$$\hat{\underline{x}}_{MAP}(\underline{y}) = \hat{\underline{x}}_{ML}(\underline{y}) = \arg\max_{\underline{x}} Q(\underline{y} \mid \underline{x}).$$

In class, we have shown that when the crossover probability p of the BSC is < 0.5, the ML decoder is the same as the minimum distance decoder.

So, we will use minimum-distance decoder here:

$$\hat{\underline{x}}(\underline{y}) = \arg\min_{\underline{x}} d(\underline{x}, \underline{y})$$

01001

$\underline{x}$	$d(\underline{x}, \underline{y})$	
00000	2	
01000	1	← minimum
10001	2	
11111	3	

From the table, we see that the most likely transmitted codeword is 01000.

(b)

$d_{min} = 1$

$\underline{x}$	00000	01000	10001	11111
00000		1	2	5
01000			3	4
10001				3
11111				

(c) For transmission of block codes over BSC, we have

$$\hat{\underline{x}}_{MAP}(\underline{y}) = \arg\max_{\underline{x}} \left(\frac{p}{1-p}\right)^{d(\underline{x},\underline{y})} p(\underline{x})$$

Here, we have $p = 0.1$. So, $\frac{p}{1-p} = \frac{0.1}{0.9} = \frac{1}{9}$

01001

$\underline{x}$	$d(\underline{x},\underline{y})$	$p(\underline{x})$	$\left(\frac{p}{1-p}\right)^{d(\underline{x},\underline{y})} p(\underline{x})$
00000	2	0.1	0.0012
01000	1	0.1	0.0111 ← maximum
10001	2	0.1	0.0012
11111	3	0.7	0.0010

From the table, we see that the most likely transmitted codeword is still 01000.

Q4: H and I

Wednesday, October 02, 2013 10:19 AM

First, we need to find the unknown constant c.

The given description for the joint pmf can be expressed using the joint pmf matrix as

$$P = \begin{array}{c|cc} x \backslash y & 2 & 4 \\ \hline 1 & 3c & 5c \\ 3 & 5c & 7c \end{array}$$

Recall that, for any joint pmf, we must have $\sum_x \sum_y \underbrace{p(x,y)}_{P[X=x,Y=y]} = 1.$

Therefore, the sum of all elements in the P matrix should be 1:

Here, we have $3c + 5c + 5c + 7c = 1 \Rightarrow 20c = 1 \Rightarrow c = \frac{1}{20}.$

a) $H(X,Y) = H\left(\underbrace{\left[\frac{3}{20}\ \frac{1}{4}\ \frac{1}{4}\ \frac{7}{20}\right]}_{\text{all elements in the P matrix}}\right) = -\frac{3}{20}\log_2\frac{3}{20} - \frac{2}{4}\log_2\frac{1}{4} - \frac{7}{20}\log_2\frac{7}{20}$

≈ 1.9406 bits.

To find $H(X)$ and $H(Y)$, we need the marginal pmfs $p(x)$ and $q(y)$, respectively.

These can be found from the sums along the rows and columns of the P matrix.

$$\begin{array}{c|cc|l} x \backslash y & 2 & 4 & p(x) \\ \hline 1 & 3c & 5c & \rightarrow 8c = \frac{8}{20} = \frac{2}{5} \\ 3 & 5c & 7c & \rightarrow 12c = \frac{12}{20} = \frac{3}{5} \\ \hline & \downarrow & \downarrow & \\ & 8c & 12c & \\ & \shortparallel & \shortparallel & \\ q(y) & 2/5 & 3/5 & \end{array}$$

b) $H(X) = H\left(\left[\frac{2}{5}\ \frac{3}{5}\right]\right) \approx 0.9710$

c) $H(Y) = H\left(\left[\frac{2}{5}\ \frac{3}{5}\right]\right) \approx 0.9710$

d) $H(X|Y) = H(X,Y) - H(Y) \approx 0.9697$

e) $H(Y|X) = H(X,Y) - H(X) \approx 0.9697$

f) $I(X;Y) = H(X) + H(Y) - H(X,Y) \approx 0.0013$

Q5: H and I (when X and Y are independent)

Wednesday, October 02, 2013 10:19 AM

First, we need to find the unknown constant β

The given description for the joint pmf can be expressed using the joint pmf matrix P as

$$P = \begin{array}{c|cc} x \backslash y & 1 & 3 \\ \hline 3 & 1/15 & 4/15 \\ 4 & 2/15 & \beta \end{array}$$

Recall that, for any joint pmf, we must have $\sum_x \sum_y \overbrace{p(x,y)}^{P[X=x,Y=y]} = 1$.

Therefore, the sum of all elements in the P matrix should be 1:

Here, we have $\frac{1}{15} + \frac{4}{15} + \frac{2}{15} + \beta = 1 \Rightarrow \beta = 1 - \frac{7}{15} = \frac{8}{15}$.

(a) To check independence, we need the marginal pmfs $p(x)$ and $q(y)$, respectively.

These can be found from the sums along the rows and columns of the P matrix.

$$P = \begin{array}{c|cc|c} x \backslash y & 1 & 3 & p(x) \\ \hline 3 & 1/15 & 4/15 & \to 5/15 = 1/3 \\ 4 & 2/15 & 8/15 & \to 10/15 = 2/3 \\ \hline & \downarrow & \downarrow & \\ & 3/15 & 12/15 & \\ q(y) & = 1/5 & = 4/5 & \end{array}$$

$\Rightarrow \underline{p} = [1/3 \quad 2/3]$

$\Rightarrow \underline{q} = [1/5 \quad 4/5]$

Recall that two random variables X and Y are independent if and only if

$p(x,y) = p(x)q(y)$ for all pair (x,y)

This is equivalent to $P = \underline{p}^T \underline{q}$.

Notice that in this problem, $P = \underline{p}^T \underline{q}$. Therefore, $X \perp\!\!\!\perp Y$.

(i) $H(X,Y) \overset{X \perp\!\!\!\perp Y}{=} H(X) + H(Y) \approx 1.6402$

(ii) $H(X) \overset{X \perp\!\!\!\perp Y}{=} H([1/3 \quad 2/3]) \approx 0.9183$

(iii) $H(Y) \overset{X \perp\!\!\!\perp Y}{=} H([1/5 \quad 4/5]) \approx 0.7219$

(iv) $H(X|Y) = H(X) \approx 0.9183$

(v) $H(Y|X) \overset{X \perp\!\!\!\perp Y}{=} H(Y) \approx 0.7219$

(vi) $I(X;Y) = 0$ because $X \perp\!\!\!\perp Y$