Information and Coding Theory. Homework 1

Andrei Dzis

February 2019

1 Problem

1.1 Problem Statement

(1 point) A fair coin is flipped until the first head occurs. Let's X denote the number of flips required. Find the entropy $H(X)$ in bits.

1.2 Solution

First of all, as it was mentioned, we have fair coin, that mean that probability of head and tail are equal, $\mathbb{P}(\text{head}) = \mathbb{P}(\text{tail}) = \frac{1}{2}$. So if X is the number of flips required until the first head occurs, so previous flips tails occur. So random variable X from previous thoughts can be described in such way:

$$X : \mathscr{X} = \{1, 2, 3, \dots\}, P_x = \{\frac{1}{2}, (\frac{1}{2})^2, (\frac{1}{2})^3, \dots\} \tag{1}$$

So probability $\mathbb{P}(X = i) = (\frac{1}{2})^i$ So we can count the entropy:

$$H(X) = -\sum_{i=1}^{\infty} \mathbb{P}(X = i) \log(\mathbb{P}(X = i)) = -\sum_{i=1}^{\infty} (\frac{1}{2})^i i \log(\frac{1}{2}) = \sum_{i=1}^{\infty} \frac{i}{2^i} = 2 \tag{2}$$

The last equation is done using that fact, that $\sum\limits_{i=1}^{\infty} i(k)^i = \frac{k}{(1-k)^2}$.
So the resulting entropy equals 2 bits.

2 Problem

2.1 Problem Statement

(1 point) Let $p = (p_1, \dots, p_n)$ be a probability vector, i.e $p_i \geq 0, i = 1, \dots, n$ and $\sum_{i=1}^{n} p_i = 1$
(a) What is the maximal value of $H(p)$ when p ranges over the set of n-dimensional probability vectors? What is the optimal distribution?
(b) What is the maximal value of $H(p)$ when p ranges over the set of n-dimensional probability vectors, such that $p_1 = \alpha, 0 \leq \alpha \leq 1$? What is the optimal distribution?

2.2 Solution

- Let's solve conditional extremum problem using Lagrange multiplier:

$$L(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log(p_i) + \lambda(\sum_{i=1}^{n} p_i - 1) \tag{3}$$

  Let's differentiate with respect to p_i:

$$-\log(p_i) - \frac{1}{\ln 2} + \lambda = 0 \tag{4}$$

$$-\ln(p_i) = 1 - \lambda \ln 2 \to p_i = e^{\lambda \ln 2 = 1} \tag{5}$$

  Using the fact that $\sum\limits_{i=1}^{n} p_i = 1$ we receive from 5:

$$ne^{\lambda \ln 2 = 1} = 1 \to \ln(\frac{1}{n}) = \lambda \ln 2 - 1 \to \lambda = \frac{1 + \ln(\frac{1}{n})}{\ln 2} \tag{6}$$

  Substitute λ into the original equation for p_i:

$$-\frac{\ln(p_i)}{\ln 2} - \frac{1}{\ln 2} + \frac{1 + \ln(\frac{1}{n})}{\ln 2} = 0 \to \ln(p_i) = \ln(\frac{1}{n}) \to p_i = \frac{1}{n} \tag{7}$$

  This is the optimal distribution.
  Now let's count entropy:

$$H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i = -n * \frac{1}{n} \log(\frac{1}{n}) = -\log(\frac{1}{n}) \tag{8}$$

- Now we have fixed value for $p_1 = \alpha$, so we can write down conditional extremum problem, using that fact that:

$$\sum_{i=1}^{n} p_i = \alpha + \sum_{i=2}^{n} p_i = 1 \to \sum_{i=2}^{n} p_i = 1 - \alpha$$

$$L(\mathbf{p}) = -\alpha \log(\alpha) - \sum_{i=2}^{n} p_i \log(p_i) + \lambda(\sum_{i=2}^{n} p_i - 1 + \alpha) \tag{9}$$

  Let's differentiate with respect to p_i:

$$-\log(p_i) - \frac{1}{\ln 2} + \lambda = 0 \to \log(p_i) = \lambda - \frac{1}{\ln 2} \to p_i = e^{\lambda - \frac{1}{\ln 2}} \tag{10}$$

  Summing last equation from $i = 2$ to n:

$$(n-1)e^{\lambda - \frac{1}{\ln 2}} = 1 - \alpha \to \lambda = \ln(\frac{1-\alpha}{1-n}) + \frac{1}{\ln(2)} \tag{11}$$

So putting the last result into equation for p_i:

$$\log(p_i) = \log(\frac{1-\alpha}{1-n}) \to p_i = \frac{1-\alpha}{1-n}, \text{for } i = 2, \dots, n \tag{12}$$

This is the optimal distribution.
Let's calculate the entropy:

$$H(\mathbf{p}) = -\alpha \log(\alpha) - (1-\alpha)\log(\frac{1-\alpha}{n-1}) \tag{13}$$

3 Problem

3.1 Problem Statement

(1 point) Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

(a) Find a binary Huffman code for X. Find the expected code length for this encoding.
(b) Find a ternary Huffman code for X. Find the expected code length for this encoding.

3.2 Solution

Let's build binary Huffman code, results are represented in a table:

x_i	Codeword	l_i
x_1	1	1
x_2	00	2
x_3	011	3
x_4	01000	5
x_5	01001	5
x_6	01010	5
x_7	01011	5

Let's count the expected code length for this encoding:

$$L(x) = \sum_{i=1}^{7} p_i * l_i = 1*0.49+2*0.26+3*0.12+5*0.04+5*0.04+5*0.03+5*0.02 =$$

$$= 2.02 \text{ bits} \tag{14}$$

Let's build ternary Huffman code, results are represented in a table:

x_i	Codeword	l_i
x_1	0	1
x_2	1	1
x_3	20	2
x_4	22	2
x_5	210	3
x_6	211	3
x_7	212	3

Let's count the expected code length for this encoding:

$$L(x) = \sum_{i=1}^{7} p_i * l_i = 1*0.49+1*0.26+2*0.12+2*0.04+3*0.04+3*0.03+3*0.02 = \\ = 1.34 \text{ bits} \tag{15}$$

4 Problem

4.1 Problem Statement

(1 point) Consider a discrete memoryless channel (DMC) with $\mathscr{X} = \mathscr{Y} = \{0, 1\}$ and a probability transition matrix

$$P_{Y|X} = \begin{pmatrix} 1 & 0 \\ \sigma & 1-\sigma \end{pmatrix}$$

where rows and columns correspond to elements of $\mathscr{X}$ and $\mathscr{Y}$ accordingly.

(a) Find the capacity of the channel.

(b) Find the limit of capacity and the capacity achieving distributions P_X when $\sigma \to 1$

4.2 Solution

$$P_{Y|X} = \begin{pmatrix} 1 & 0 \\ \sigma & 1-\sigma \end{pmatrix} = \begin{pmatrix} P_{Y|X}(Y=0|X=0) & P_{Y|X}(Y=1|X=0) \\ P_{Y|X}(Y=0|X=1) & P_{Y|X}(Y=1|X=1) \end{pmatrix}$$

For capacity we can write down:

$$C = \max_{P_x}\{I(X,Y)\} \tag{16}$$

$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{x \in \mathbf{X}} P(X=x) H(Y|X=x) = \tag{17}$$

Due to the fact that $0\log(0) = 1\log(1) = 0$:

$$= H(Y) + P(X=1)(\sigma\log(\sigma) + (1-\sigma)\log(1-\sigma)) = H(Y) - P(X=1)h(\sigma) \tag{18}$$

For $P(Y=y)$, denoting $P(X=1)=\pi$, then $P(X=0)=1-P(X=1)=1-\pi$:

$$P(Y=0)=\sum_{x\in\mathbf{X}} P(X=x)P_{Y|X}(Y=0|X=x)=1-\pi+\sigma\pi \tag{19}$$

Then:

$$P(Y=1)=(1-\sigma)\pi \tag{20}$$

Now for $H(Y)$ we can write:

$$H(Y)=-(1-\pi+\sigma\pi)\log(1-\pi+\sigma\pi)-((1-\sigma)\pi)\log((1-\sigma)\pi) \tag{21}$$

So collecting all needed results we can write:

$$I(X,Y)=-(1-\pi+\sigma\pi)\log(1-\pi+\sigma\pi)-((1-\sigma)\pi)\log((1-\sigma)\pi)-\pi h(\sigma) \tag{22}$$

$$I(X,Y)=-((\sigma-1)\pi+1)\log((\sigma-1)\pi+1)-((1-\sigma)\pi)\log((1-\sigma)\pi)-\pi h(\sigma) \tag{23}$$

Let's differentiate with respect to π:

$$(1-\sigma)\log((\sigma-1)\pi+1)-\frac{\sigma-1}{\ln 2}-(1-\sigma)\log((1-\sigma)\pi)-\frac{1-\sigma}{\ln 2}-h(\sigma)=0 \tag{24}$$

$$(1-\sigma)\log(\frac{\pi(\sigma-1)+1}{(1=\sigma)\pi})=h(\sigma) \tag{25}$$

$$\frac{h(\sigma)}{1-\sigma}=\log(\frac{\pi(\sigma-1)+1}{(1-\sigma)\pi})\rightarrow 2^{\frac{h(\sigma)}{1-\sigma}}=\frac{\pi(\sigma-1)+1}{(1-\sigma)\pi} \tag{26}$$

$$\pi=\frac{1}{(1-\sigma)(2^{\frac{h(\sigma)}{1-\sigma}}+1)} \tag{27}$$

Now we can put into the capacity the resulted π:

$$I(X,Y)=-\frac{2^{\frac{h(\sigma)}{1-\sigma}}}{2^{\frac{h(\sigma)}{1-\sigma}}+1}\log(\frac{2^{\frac{h(\sigma)}{1-\sigma}}}{2^{\frac{h(\sigma)}{1-\sigma}}+1})-\frac{1}{(2^{\frac{h(\sigma)}{1-\sigma}}+1)}\log(\frac{1}{(2^{\frac{h(\sigma)}{1-\sigma}}+1)})-\frac{h(\sigma)}{(1-\sigma)(2^{\frac{h(\sigma)}{1-\sigma}}+1)} \tag{28}$$

$$I(X,Y)=\log(2^{\frac{h(\sigma)}{1-\sigma}}+1))-\frac{h(\sigma)}{1-\sigma}\frac{2^{\frac{h(\sigma)}{1-\sigma}}}{2^{\frac{h(\sigma)}{1-\sigma}}+1}-\frac{h(\sigma)}{(1-\sigma)(2^{\frac{h(\sigma)}{1-\sigma}}+1)} \tag{29}$$

$$C=\max_{P_x}\{I(X,Y)\}=\log(2^{\frac{h(\sigma)}{1-\sigma}}+1))-\frac{h(\sigma)}{1-\sigma} \tag{30}$$

Now let's calculate the limit of π, C, when $\sigma\rightarrow 1$

$$\lim_{\sigma\rightarrow 1}\frac{1}{(1-\sigma)(2^{\frac{h(\sigma)}{1-\sigma}}+1)}=|\text{denote } \sigma-1=x|$$

$$=\lim_{x\rightarrow 0}\frac{1}{x(2^{\frac{h(x-1)}{x}}+1)}=\lim_{x\rightarrow 0}2^{-\log(x)+\frac{1-x}{x}\log(1-x)+\frac{x}{x}\log(x)}=$$

$$= \lim_{x\to 0} 2^{\frac{(1-x)(-x-\frac{x^2}{2}-\frac{x^3}{3!}-\dots)}{x\ln 2}} = 2^{-\frac{1}{\ln 2}} \tag{31}$$

The limit for C we can find from 23:

$$\lim_{\sigma\to 1} C = \lim_{\sigma\to 1}\Big(-((\sigma-1)\pi+1)\log((\sigma-1)\pi+1)-((1-\sigma)\pi)\log((1-\sigma)\pi)-\pi h(\sigma)\Big) =$$

$$= -1\log 1 - 0\log 0 - (1) = -\pi(-1\log 1 - 0\log 0) = 0 \tag{32}$$

5 Problem

5.1 Problem Statement

(1 point) Consider a binary symmetric channel (BSC) with transition probability p. The output ofthis channel is fed to the input of a binary erasure channel (BEC) with erasure probability ϵ. What is the capacity of the resulting channel?

5.2 Solution

For BSC we have for capacity:

$$C_{BSC} = 1 - h(p) \tag{33}$$

For BEC we have for capacity:

$$C_{BEC} = 1 - \epsilon \tag{34}$$

Let's denote $\mathcal{X} \in \{0,1\}$ as the input for BSC and as $\mathcal{Y} \in \{0,1,\epsilon\}$ we denote the output of the whole system. Also let's denote $\mathbb{P}(X=1)=\pi$, then $\mathbb{P}(X=0)=1-\pi$. So we can write:

$$\mathbb{P}(y=0) = (1-\epsilon)((1-\pi)(1-p)+\pi p) \tag{35}$$

$$\mathbb{P}(y=1) = (1-\epsilon)(\pi(p-1)+(1-\pi)p) \tag{36}$$

$$\mathbb{P}(y=\epsilon) = \epsilon \tag{37}$$

Then we can write:

$$H(Y) = -\sum_{y\in\mathcal{Y}} \mathbb{P}(Y=y)\log(\mathbb{P}(Y=y)) \tag{38}$$

$$H(Y) = -(1-\epsilon)((1-\pi)(1-p)+\pi p)\log((1-\epsilon)((1-\pi)(1-p)+\pi p)) - \epsilon\log(\epsilon) - (1-\epsilon)(\pi(p-1)+(1-\pi)p)\log((1-\epsilon)(\pi(p-1)+(1-\pi)p))$$

Let's write for conditional entropy:

$$H(Y|X) = \sum_{x\in\mathcal{X}} \mathbb{P}(X=x)H(Y|X=x) \tag{39}$$

$$H(Y|X=0) = -(1-p)(1-\epsilon)\log((1-p)(1-\epsilon)) - \epsilon\log(\epsilon) - p(1-\epsilon)\log(p(1-\epsilon)) \tag{40}$$

For $H(Y|X=1)$:

$$H(Y|X=1) = -(1-p)(1-\epsilon)\log((1-p)(1-\epsilon)) - \epsilon\log(\epsilon) - p(1-\epsilon)\log(p(1-\epsilon)) \tag{41}$$

So, after putting 40 and 41 into 39, we receive that $H(Y|X) = H(Y|X=0) = H(Y|X=1)$. So the final moment is to find a capacity for our system of channels, it can be done using formula:

$$C_{system} = \max\{I(Y,X)\} = \max\{H(Y) - H(Y|X)\} = \max\{H(Y)\} - H(Y|X) \tag{42}$$

So in order maximize the mutual information, we need to maximize $H(Y)$. So we need to differentiate $H(Y)$ with respect to π:

$$\frac{\partial H(Y)}{\partial(\pi)} = (1-\epsilon)(2p-1)\log((1-\epsilon)((1-\pi)(1-p)+\pi p)) + \tag{43}$$

$$+\frac{(1-\epsilon)(2p-1)}{\ln 2} + (1-\epsilon)(1-2p)\log((1-\epsilon)(\pi(1-p)+(1-\pi)p)) + \frac{(1-\epsilon(1-2p))}{\ln 2} = 0 \tag{44}$$

From this equation we receive maximum condition:

$$(1-\epsilon)((1-\pi)(1-p)+\pi p) = (1-\epsilon)(\pi(1-p)+(1-\pi)p) \to \pi = \frac{1}{2} \tag{45}$$

So let's put received value for π into capacity:

$$C_{system} = -(1-\epsilon)log(\frac{1}{2}(1-\epsilon)) - \epsilon\log(\epsilon) - (1-p)(1-\epsilon)\log((1-p)(1-\epsilon)) + \epsilon\log(\epsilon) + \tag{46}$$

$$+p(1-\epsilon)\log(p(1-\epsilon)) = (1-\epsilon)(1-p\log p + (1-p)\log(1-p)) = C_{BEC}C_{BSC} \tag{47}$$

So the capacity of our system is a product of system parts' capacities.

6 Problem

6.1 Problem Statement

(1 point) Find the capacity of 6 parallel independent discrete time Gaussian channels, such a noise variance have the following values $N_1 = N_2 = N_3 = \mathbb{E}[Z^2] = 1, N_4 = \mathbb{E}[Z^2] = 2, N_5 = N_6 = \mathbb{E}[Z^2] = 3$. The total power constraint is equal to $P = 11$.

6.2 Solution

To find maximum capacity, distributing total power among 6 channels we should use waterfilling technique. Let's denote by $P_i, i = 1, \ldots, 6$ the power distributed for i-th channel. So the problem is reduced to finding the power allotment that maximizes the capacity subject to the constraint that $\sum_{i=1}^{6} P_i = P = 11$. For every channel we can write:

$$P_i = (\nu - N_i)^+, \tag{48}$$

where:

$$(X)^+ = \begin{cases} X & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases} \tag{49}$$

So we can check easily, that if we have $\nu \leq \max\{N_i\}$ we have the total power less then $P = 11$. So we can rewrite 48 using this note:

$$P_i = \nu - N_i \tag{50}$$

So let's sum for every i:

$$\sum_{i=1}^{6} P_i = 6\nu - \sum_{i=1}^{6} N_i \to \nu = \frac{22}{6} = 3\frac{2}{3} \tag{51}$$

Now, let's calculate each power for each channel separately:

$$P_1 = P_2 = P_3 = 2\frac{2}{3} \tag{52}$$

$$P_4 = 1\frac{2}{3} \tag{53}$$

$$P_5 = P_6 = \frac{2}{3} \tag{54}$$

As we can see $P = \sum_{i=1}^{6} P_i = 3 * 2\frac{2}{3} + 1\frac{2}{3} + 2 * \frac{2}{3} = 11$ Because of the fact that $I(X_1, \ldots, X_6, Y_1, \ldots, Y_6) \leq \sum_i \frac{1}{2} \log(1 + \frac{P_i}{N_i})$, we can calculate capacity by following:

$$C = \sum_i \frac{1}{2} \log(1 + \frac{P_i}{N_i}) = 3 * \frac{1}{2} \log(1 + 2\frac{2}{3}) + \frac{1}{2} \log(1 + \frac{5}{6}) + 2 * \frac{1}{2} \log(1 + \frac{2}{9}) \tag{55}$$

So:

$$C = 3.538 \text{ bits/s} \tag{56}$$

7 Problem

7.1 Problem Statement

(2 points) Let two users use the same alphabet, i.e. $\mathbf{X}_1 = \mathbf{X}_2 = \{-1, 1\}$

(a) Find the capacity region of noiseless adder multiple access channel $Y = X_1 + X_2$

(b) Assume we want to construct uniquely decodable codebook C of length n and size 2^{Rn} for this channel. I.e. we require the sums c_1+c_2, where $c_1, c_2 \in C$, to be all different. Please give an upper bound on the rate of such codebook and compare it to the capacity when $n \to \infty$. Why do you see a big difference?

7.2 Solution

For two users capacity region can be described as follows:

$$R_1 \leq I(X_1, Y|X_2) \tag{57}$$

$$R_2 \leq I(X_2, Y|X_1) \tag{58}$$

$$R_1 + R_2 \leq I(X_1, X_2|Y) \tag{59}$$

R_1 is maximum, when 2-nd user sends no information, but chooses a fixed input, to maximize the mutual information, and vice versa for 2-nd user. For this situation we can send at rate 1, thus:

$$R_1^{max} = R_2^{max} = 1 \tag{60}$$

Then:

$$R_1 \leq 1, R_2 \leq 1 \tag{61}$$

For sum of rates:

$$R_1 + R_2 \leq I(X_1, X_2|Y) = H(Y) - H(Y|X_1, X_2) = H(Y) \tag{62}$$

Assuming $\mathbb{P}(x_1 = -1) = \mathbb{P}(x_2 = -1) = \frac{1}{2}$, we receive, that $\mathbb{P}(Y = -2) = \mathbb{P}(Y = 2) = \frac{1}{4}$ and hence $H(Y) = 1.5$ That happens, when transmitter of X_1 sends with probability $\mathbb{P}(x_1 = -1)$. Hence, $I(X_1, Y|X_2) = H(X_1) - H(X_1|Y, X_2) = 1$. For the other user the channel looks like an erasure channel with erasure probability $\frac{1}{2}$.

Let's calculate the number of all different word pairs $(c_1, c_2) \in \mathcal{C}$, it's equal:

$$N = |\text{number of pairs}| = 2^{2Rn-1} \tag{63}$$

For creating unique decodable code we need all pairs of (c_1, c_2) from codebook $\mathcal{C}$ to have different sums $c_1 + c_2, (c_1, c_2) \in \mathcal{C}$, the maximal number of this sums:

$$N = |\text{number of sums}| = 3^n \tag{64}$$

From this equations we have:

$$2^{2Rn-1} \leq 3^n \tag{65}$$

So for rate:

$$R \leq \log_4(3) + \frac{1}{2n} \tag{66}$$

So for sum:

$$R_{12} \leq \log_4 3 + \log_4 3 + \frac{1}{n} \rightarrow R_{12} \leq \log_2(3) + \frac{1}{n} \tag{67}$$

When $n \to \infty$, $R_{12} = \log_2(3) \approx 1.58$. This value is bigger, than the capacity, but the case is that in real life problems there's no guarantee of finding this set of code vectors, which guarantee number $N = 3^n$ of different sums of their different pairs.

8 Problem

8.1 Problem Statement

(2 points) Let us consider a channel W with real input X and binary output Y, such that

$$Pr(Y = 0|X = x) = 1 - x$$

and

$$Pr(Y = 1|X = x) = x$$

Also we have an additional constraint: $E[|X - 1/2|] \leq P$

(a) What is the capacity of this channel?

(b) Is there a simple coding scheme achieving this performance?

8.2 Solution

Let's find the range for variable x. This comes from the fact, that probability P must satisfy:

$$0 \leq P \leq 1 \rightarrow P \in [0, 1] \tag{68}$$

From this fact we conclude, that:

$$\mathbb{P}(Y = 1|X = x) \in [0, 1] \rightarrow \mathbb{P}(Y = 1|X = x) = x \rightarrow x \in [0, 1] \tag{69}$$

For mutual information, we can write:

$$I(X, Y) = H(Y) - H(Y|X) \tag{70}$$

For expected value (will be denoted as $\mathbb{E}(X)$) we can write:

$$\mathbb{E}(X) = \int_0^1 xp(x)dx \tag{71}$$

Thus, we have:

$$\mathbb{P}(Y=1)=\int_0^1 xp(x)dx=\mathbb{E} \tag{72}$$

From this fact, we can rewrite $H(Y)$ as follows:

$$H(Y)=-\mathbb{P}(Y=0)\log(\mathbb{P}(Y=0))-\mathbb{P}(Y=1)\log(\mathbb{P}(Y=1))=h(\mathbb{E}(X)) \tag{73}$$

Also, for $H(Y|X)$ we can write down:

$$H(Y|X)=\int_0^1 h(x)p(x)dx=\mathbb{E}(h(X)) \tag{74}$$

Since, for capacity we can write down:

$$C=\max\{I(X,Y)\}=\max\{h(\mathbb{E}(X))-\mathbb{E}(h(X))\} \tag{75}$$

subject to

$$\mathbb{E}(|X-\frac{1}{2}|)\leq P \tag{76}$$

Due to the fact, that function $h(X)_{max}=1$ reaches it's maximum at $X=\frac{1}{2}$, in order to maximize $h(\mathbb{E}(X)$ it's required $\mathbb{E}(X)=\frac{1}{2}$. Let's denote a distribution function of a random variable X as $F_X(x)$. For every such a function we can make it symmetric about the point $x=\frac{1}{2}$, denoting as $S_X(x)$, denoting X^* as symmetric random variable of original one about point $x=\frac{1}{2}$, with the following properties:

$$\mathbb{E}(h(X))=\mathbb{E}(h(X^*))$$

and

$$\mathbb{E}(|X-\frac{1}{2}|)=\mathbb{E}(|X^*-\frac{1}{2}|).$$

Here and below using notation symmetric means symmetric about point $x=\frac{1}{2}$. For probability density function of symmetric distribution function, we can write:

$$p_s(x)=\frac{1}{2}p(x)+\frac{1}{2}p(1-x), \tag{77}$$

where $p(x)$ is a probability density function of original distribution. For symmetric one, using that fact, that $h(x)$ is symmetric about the point $x=\frac{1}{2}$:

$$\mathbb{E}h(X^*)=\int_0^1 h(x)\frac{1}{2}p(x)+\frac{1}{2}p(1-x)dx=E(h(X)) \tag{78}$$

$$\mathbb{E}(|X^*-\frac{1}{2}|)=\int_0^{\frac{1}{2}}(\frac{1}{2}-x)(\frac{1}{2}p(x)+\frac{1}{2}p(1-x))dx+\int_{\frac{1}{2}}^1(x-\frac{1}{2})\frac{1}{2}(p(x)+\frac{1}{2}p(1-x))dx=$$

$$=\mathbb{E}(|X-\frac{1}{2}|) \tag{79}$$

So proved this properties. From our observation, the capacity is being reached, when $H(Y) - H(Y|X)$ is maximal, which in term of written above means:

$$\max(H(Y)) - \min(H(Y|X)) = \max(\mathbb{E}(h(X))) - \min(h(\mathbb{E}(X))) \tag{80}$$

We need to find such distribution function to minimize $\mathbb{E}(h(X))$ and after that make it symmetric about $x = \frac{1}{2}$, because it's the way we maximize $h(\mathbb{E}(X))$

$$\mathbb{E}(|X - \frac{1}{2}|) \leq P, P \leq \frac{1}{2} \tag{81}$$

For $\mathbb{E}(|X - \frac{1}{2}|)$:

$$\mathbb{E}(|X - \frac{1}{2}|) = \int_0^{\frac{1}{2}} h(x)p(x)dx = \frac{1}{2} - \mathbb{E}(x) \leq P \tag{82}$$

Then:

$$E(x) \geq \frac{1}{2} - P \tag{83}$$

For $\mathbb{E}(h(X))$:

$$\mathbb{E}(h(X)) = \int_0^{\frac{1}{2}} h(x)p(x)dx = 2E(x) \tag{84}$$

Thus:

$$E(h(X)) \geq 1 - 2P \tag{85}$$

Considering required distribution function as:

$$p(x) = 2P\delta(x) + (1 - 2P)\delta(x - \frac{1}{2}), \tag{86}$$

where $\delta(x)$ is the delta-function.
For this distribution $\mathbb{E}(X) = \frac{1}{2} - P$ and $\mathbb{E}(h(X)) = 1 - 2P$, so we see that is the density function, which minimizes $H(Y|X)$, for $\forall\ P \leq \frac{1}{2}$.
From previous points it was said, that after finding the density, which minimizes $H(Y|X)$, we need to make symmetrical this function about point $x = \frac{1}{2}$ to maximize $H(Y) = h(\mathbb{E}(X))$ in case that $h(\frac{1}{2}) = 1$ - maximal value.

$$p_s(x) = P\delta(x) + P\delta(x - 1) + (1 - 2P)\delta(x - \frac{1}{2}) \tag{87}$$

with the resulting capacity:

$$C = \begin{cases} 2P & \text{if } P \leq \frac{1}{2} \\ 1 & \text{if } P \geq \frac{1}{2} \end{cases}$$