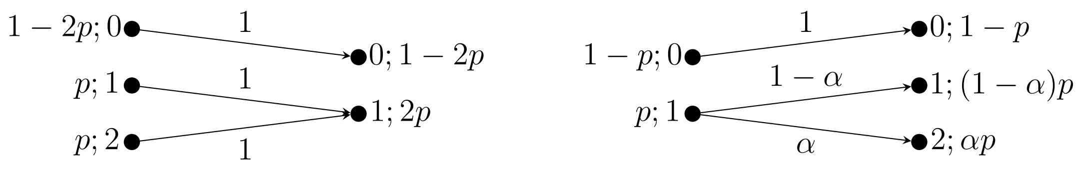
## Problem 1

The probabilities for $X$ and $Y$ for the two channels are shown in the following figure.




(A) (B)

(a) The mutual information in Figure A is

$$I(X;Y) = H(Y) - H(Y|X) = h(2p) - \big((1-2p)h(1) + ph(1) + ph(1)\big) = h(2p)$$

(b) For Figure B it is

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) = H\big(1-p, (1-\alpha)p, \alpha p\big) - \big((1-p)h(1) + ph(\alpha)\big) \\ &= -(1-p)\log(1-p) - (1-\alpha)p\log(1-\alpha)p - \alpha p \log \alpha p + p\alpha \log \alpha + p(1-\alpha)\log(1-\alpha) \\ &= h(p) \end{aligned}$$

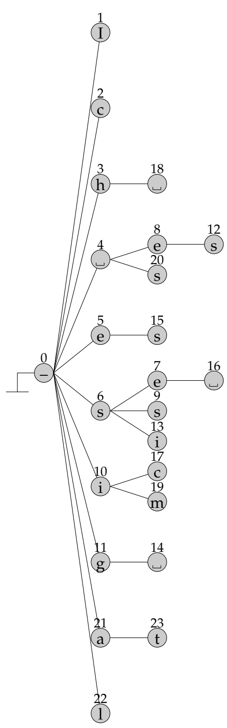
where the last inequallity is found by breaking up the logarithms in sums.

(c) A: $C = \max_p h(2p) = 1$, for $p = 1/4$
B: $C = \max_p h(p) = 1$, for $p = 1/2$

(d) In both cases the capacity is upper bounded by $C \leqslant \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\} = 1$. In channel A one bit can be transmitted by $0 \to 0$ and $\{1,2\} \to 1$. For channel B one bit can always be determined since $0 \to 0$ and $1 \to \{1,2\}$.

## Problem 2

| Index | Codeword | Dict | Bits | Binary |
|---|---|---|---|---|
| 1 | (0,I) | [I] | 0 | 01001001 |
| 2 | (0,c) | [c] | 1 | 0 01100011 |
| 3 | (0,h) | [h] | 2 | 00 01101000 |
| 4 | (0, ) | [␣] | 2 | 00 00100000 |
| 5 | (0,e) | [e] | 3 | 000 01100101 |
| 6 | (0,s) | [s] | 3 | 000 01110011 |
| 7 | (6,e) | [se] | 3 | 110 01100101 |
| 8 | (4,e) | [␣e] | 3 | 100 01100101 |
| 9 | (6,s) | [ss] | 4 | 0110 01110011 |
| 10 | (0,i) | [i] | 4 | 0000 01101001 |
| 11 | (0,g) | [g] | 4 | 0000 01100111 |
| 12 | (8,s) | [␣es] | 4 | 1000 01110011 |
| 13 | (6,i) | [si] | 4 | 0110 01101001 |
| 14 | (11, ) | [g␣] | 4 | 1011 00100000 |
| 15 | (5,s) | [es] | 4 | 0101 01110011 |
| 16 | (7, ) | [se␣] | 4 | 0111 00100000 |
| 17 | (10,c) | [ic] | 5 | 01010 01100011 |
| 18 | (3, ) | [h␣] | 5 | 00011 00100000 |
| 19 | (10,m) | [im] | 5 | 01010 01101101 |
| 20 | (4,s) | [␣s] | 5 | 00100 01110011 |
| 21 | (0,a) | [a] | 5 | 00000 01100001 |
| 22 | (0,l) | [l] | 5 | 00000 01101100 |
| 23 | (21,t) | [at] | 5 | 10101 01110100 |



Number code bits: $1 + 2 \cdot 2 + 4 \cdot 3 + 8 \cdot 4 + 7 \cdot 5 + 23 \cdot 8 = 268$
Number uncoded bits: $38 \cdot 8 = 304$
Rate: $R = 268/304 = 0.8816$

## Problem 3

Let the outcome of $X$ be W and B, for white and black respectively. Then the probabilities for $X$ conditioned on the urn, $Y$ is as in the following table. Furthermore, since the choise of urn are equally likely the joint probability is $p(x, y) = \frac{1}{2}p(x|y)$.

| $p(x\|y)$ | W | B |
|---|---|---|
| 1 | 4/7 | 3/7 |
| 2 | 3/10 | 7/10 |

| $p(x, y)$ | W | B |
|---|---|---|
| 1 | 2/7 | 3/14 |
| 2 | 3/20 | 7/20 |

(a) The distribution of $X$ is given by $P(X = \text{W}) = \frac{1}{2} \cdot \frac{4}{7} + \frac{1}{2} \cdot \frac{3}{10} = \frac{61}{140}$, and the entropy $H(X) = h\big(\frac{61}{140}\big) = 0.988$.

(b) The mutual information can be derived as

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = h\big(\tfrac{61}{140}\big) + h\big(\tfrac{1}{2}\big) - H\big(\tfrac{2}{7}, \tfrac{3}{14}, \tfrac{3}{20}, \tfrac{7}{20}\big) = 0.0548$$

(c) By adding one more urn ($Y = 3$) we get the following tables (with $p(x) = 1/3$)

| $p(x\|y)$ | W | B |
|---|---|---|
| 1 | 4/7 | 3/7 |
| 2 | 3/10 | 7/10 |
| 3 | 1 | 0 |

| $p(x, y)$ | W | B |
|---|---|---|
| 1 | 4/21 | 3/21 |
| 2 | 1/10 | 7/30 |
| 3 | 1/3 | 0 |

Hence, $P(X = \text{W}) = \frac{131}{210}$ and $P(X = \text{B}) = \frac{79}{210}$, and $H(X) = h\big(\frac{79}{210}\big)$. The mutual information is

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = h\big(\tfrac{79}{210}\big) + \log 3 - H\big(\tfrac{4}{21}, \tfrac{3}{21}, \tfrac{1}{10}, \tfrac{7}{30}, \tfrac{1}{3}\big) = 0.3331$$

## Problem 4

(a) For $P(n)$ to be a probability function it must be positive and sum to 1. Here, it is clear that $P(n) \geqslant 0$ for all $n$, and since $1/k < 1$ the sum becomes

$$\sum_{n=1}^{\infty}(k-1)k^{-1} = (k-1)\sum_{n=1}^{\infty}\big(\tfrac{1}{k}\big)n = (k-1)\frac{\frac{1}{k}}{1-\frac{1}{k}} = \frac{k-1}{k-1} = 1$$

Hence, it is a probability function.

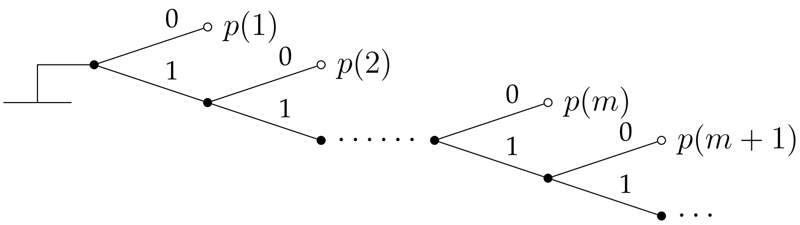
(b) With $k = 2$ we get $P(n) = \big(\frac{1}{2}\big)^n$. By considering the optimal codeword lengths

$$l_n^{(opt)} = -\log P(n) = -\log\big(\tfrac{1}{2}\big)^n = n$$

we see that this is an integer for each number $n$. It is also the same as the codeword lengths for the unary code, and we conclude that it is optimal for this case. The entropy is in that case equal to the average codeword length

$$H(X) = L = \sum_{n=1}^{\infty} n\big(\tfrac{1}{2}\big)^n = \tfrac{1}{2}\sum_{n=1}^{\infty} n\big(\tfrac{1}{2}\big)^{n-1} = \frac{\frac{1}{2}}{\big(1-\frac{1}{2}\big)^2} = 2$$

(c) For a general $k$ the optimal codeword length $l_n^{(opt)} = -\log P(n)$ is typically not integers and can therefore not be used to construct an optimal code. It also means that the average length of an optimal code will not equal the entropy. Our next attemmpt is then to show that the code satisfies Huffman's algorithm, which will produce an optimal code. Then write the code in a tree,




Consider then the sub-tree stemming from level $m$ (the tree containing the leaves $p(m+1)$, $p(m+2)$, etc. The root node of this tree has the probability

$$\begin{aligned} r(m) &= \sum_{n=m+1}^{\infty} (k-1)\big(\tfrac{1}{k}\big)^n = (k-1)\big(\tfrac{1}{k}\big)^{m+1} \sum_{l=0}^{\infty} \big(\tfrac{1}{k}\big)^l \\ &= (k-1)\big(\tfrac{1}{k}\big)^{m+1} \frac{1}{1-\frac{1}{k}} = \big(\tfrac{1}{k}\big)^m \leqslant (k-1)\big(\tfrac{1}{k}\big)^m = p(m) \end{aligned}$$

Hence, among the nodes $p(1)$, $p(2)$, …, $p(m)$ and $r(m)$, the two least probable are $p(m)$ and $r(m)$. Merging those two nodes in a tree will give one step further up in the tree. After $m-2$ more similar merges, according to the Huffman algorithm, the unary code has been constructed. Hence, for $p(n)$ as in the problem, the unary code is a Huffman code and, hence, it is optimal. The corresponding codeword length given by

$$L = \sum_{n=1}^{\infty} n(k-1)\big(\tfrac{1}{k}\big)^n = (k-1)\big(\tfrac{1}{k}\big) \sum_{n=1}^{\infty} n\big(\tfrac{1}{k}\big)^{n-1} = (k-1)\frac{\frac{1}{k}}{\big(1-\frac{1}{k}\big)^2} = \frac{k}{k-1}$$

## Problem 5

(a)

$$\begin{aligned} I(X;Y,Z) &= H(Y,Z) - H(Y,Z|X) \\ &= H(Y) + H(Z|Y) - H(Y|X) - H(Z|Y,X) \\ &= H(Y) - H(Y|X) + H(Z) - H(Z|X) - H(Z) + H(Z|Y) \\ &= I(X;Y) + I(X;Z) - I(Y;Z) \end{aligned}$$

where in the third equality the terms $H(Z) - H(Z)$ are added, and it is noted that $H(Z|Y,X) = H(Z|X)$ since the two channels work independently.

(b) Since $X$ is binary with equal probabilities we get directly $I(X;Y) = I(X;Z) = 1-h(p)$. It also gives that $p(y) = p(z) = 1/2$, and, hence, $I(Y;Z) = H(Y) + H(Z) - H(Y,Z) = 2 - H(Y,Z)$. Then, to get the first part of the problem,

$$\begin{aligned} I(X;Y,Z) &= I(X;Y) + I(X;Z) - I(Y;Z) \\ &= 2\big(1-h(p)\big) - \big(2 - H(Y,Z)\big) = H(Y,Z) - 2h(p) \end{aligned}$$

To get the distrubiution for $(Y,Z)$ we follow the hint in the problem and derive $p(y,z|x) = p(y|x)p(z|x)$, which follows from that conditioned on $X$, $Y$ and $Z$ are independent. Since $p(x) = 1/2$ the unconditional probability is $p(y,z) = \frac{1}{2}(p(y,z|x=0) + p(y,z|x=$

1)). The probability functions are listed in the following table

| $X$ | $Y$ | $Z$ | $p(y,z\|x)$ |
|---|---|---|---|
| 0 | 0 | 0 | $(1-p)^2$ |
| 0 | 0 | 1 | $p(1-p)$ |
| 0 | 1 | 0 | $p(1-p)$ |
| 0 | 1 | 1 | $p^2$ |
| 1 | 0 | 0 | $p^2$ |
| 1 | 0 | 1 | $p(1-p)$ |
| 1 | 1 | 0 | $p(1-p)$ |
| 1 | 1 | 1 | $(1-p)^2$ |

| $Y$ | $Z$ | $p(y,z)$ |
|---|---|---|
| 0 | 0 | $\frac{1}{2}(p^2+(1-p)^2)$ |
| 0 | 1 | $p(1-p)$ |
| 1 | 0 | $p(1-p)$ |
| 1 | 1 | $\frac{1}{2}(p^2+(1-p)^2)$ |

Then,

$$\begin{aligned} H(Y,Z) &= H\Big(\tfrac{1}{2}\big((1-p)^2+p^2\big), \tfrac{1}{2}\big((1-p)^2+p^2\big), p(1-p), p(1-p)\Big) \\ &= -\big((1-p)^2+p^2\big)\log\frac{(1-p)^2+p^2}{2} - 2(1-p)\log p(1-p) \end{aligned}$$

Inserting in the above expression gives

$$\begin{aligned} I(X;Y,Z) &= H(Y,Z) - 2h(p) \\ &= p^2\log\frac{2}{(1-p)^2+p^2} + (1-p)^2\log\frac{2}{(1-p)^2+p^2} \\ &\quad - 2p(1-p)\log p - 2p(1-p)\log(1-p) + 2p\log p + 2(1-p)\log(1-p) \\ &= p^2\log\frac{2}{(1-p)^2+p^2} + (1-p)^2\log\frac{2}{(1-p)^2+p^2} + p^2\log p + (1-p)^2\log(1-p) \\ &\overset{(a)}{=} p^2\log\frac{2p^2}{(1-p)^2+p^2} + (1-p)^2\log\frac{2(1-p)^2}{(1-p)^2+p^2} \\ &= \big(p^2+(1-p)^2\big) + p^2\log\frac{p^2}{(1-p)^2+p^2} + (1-p)^2\log\frac{(1-p)^2}{(1-p)^2+p^2} \\ &= \big(p^2+(1-p)^2\big)\Big(1 + \frac{p^2}{p^2+(1-p)^2}\log\frac{p^2}{(1-p)^2+p^2} + \frac{(1-p)^2}{p^2+(1-p)^2}\log\frac{(1-p)^2}{(1-p)^2+p^2}\Big) \\ &\overset{(b)}{=} \big(p^2+(1-p)^2\big)\Big(1 - h\Big(\frac{p^2}{(1-p)^2+p^2}\Big)\Big) \end{aligned}$$

where (a) and (b) are the results to be shown.

The formula in (b) can be interpreted as follows. Viewed from the receiver $(Y,Z) = (0,1)$ or $(Y,Z) = (1,0)$, which happens with probability $2p(1-p)$, the probability for $X = 0$ and $X = 1$ are both $1/2$, so there is no information in this event. On the other hand, with probability $p^2 + (1-p)^2$ the receiver gets $(0,0)$ or $(1,1)$, which gives the information $1 - h\big(\frac{p^2}{(1-p)^2+p^2}\big)$. Here, $\frac{p^2}{(1-p)^2+p^2}$ is $P(Y \neq X, Z \neq X | Y = Z)$, that is, the probability that both $Y$ and $Z$ are wrong if the receiver gets the same result from the two channels.