Montogue

Quiz EL403
Information Theory and Coding

Lucas Monteiro Nogueira

PROBLEMS

Problem 1

Regarding information and coding theory, true or false?

1.( ) Let X be a random variable taking on a finite number of values. If Y is a random variable such that $Y = 2^X$, it can be shown that $H(X) = H(Y)$ for all values of X, where $H(\cdot)$ denotes the entropy function.

2.( ) The random variable

$$X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases}$$

has entropy function $H(X)$, which can be expressed as a function $H(p)$ of probability p. The plot of $H(p)$ versus p is concave upward.

3.( ) Let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the binary entropy function. It can be shown that the average entropy when the probability p is chosen uniformly in the range $0 \leq p \leq 1$, is 0.5 bits.

4.( ) The entropy of a real-valued, discrete random variable cannot be infinite.

5.( ) If X is a Gaussian random variable with mean zero and variance 10, the differential entropy of X is greater than 3.8 bits.

6.( ) The capacity of a binary symmetric channel (BSC) with bit-error probability equal to 0.01 is greater than 0.91 bits.

7.( ) A code is a fixed-length code if all code words in the code have the same length. The American Standard Code for Information Exchange (ASCII) is an example of fixed-length code.

8.( ) It is impossible to construct a five-code word binary prefix-free code with code word lengths $\ell \in \{2,2,2,3,4\}$.

Problem 2 (Translated from Mildenberger, 1992)

A source X sends the symbols x_1, x_2, x_3, x_4. We have the probabilities $P(x_1) = 1/2$ and $P(x_2) = 1/4$. What should the probabilities $P(x_3)$ and $P(x_4)$ be to maximize the entropy of the source X? What is this maximum entropy?

Problem 3 (Translated from Mildenberger, 1992)

A composite source consists of two subsources X and Y, both of which have the character set $\{a, b, c\}$. The nine joint probabilities are compiled in the following table:

Probabilities		y_i		
		a	***b***	***c***
x_i	***a***	0	4/15	1/15
	b	8/27	8/27	0
	c	1/27	4/135	1/135

Show that the joint entropy $H(X,Y)$ is less than the sum of the individual entropies $H(X)$ and $H(Y)$. Why is this the case?

Problem 4 (Translated from Mildenberger, 1992)

Consider two independent sources X and Y with the same symbol set (a, b, c). Source X is associated with probabilities $P(x = a) = 0.2$, $P(x = b) = 0.3$, and $P(x = c) = 0.5$, while source Y has entropy $H(Y) = \log_2 3 = 1.585$ bit. Find the joint entropy $H(X,Y)$ and each of the joint probabilities $P(x_i, y_j)$.

Problem 5 (Translated from Mildenberger, 1992)

A composite source made up of two sources *X* and *Y*. The probabilities associated with source *Y* are given in continuation,

$$Y = \begin{Bmatrix} a & b & c & d \\ 1/4, & 1/4, & 1/4, & 1/4 \end{Bmatrix}$$

The joint probability matrix is only partially known; missing data are denoted with a ×.

Probabilities		y_i			
		a	***b***	***c***	***d***
x_i	***a***	1/8	1/8	1/8	1/8
	b	×	0	×	×
	c	×	×	×	1/8
	d	1/8	0	1/8	×

Find the conditional entropies *H(X|Y)* and *H(Y|X)*.

Problem 6 (Translated from Mildenberger, 1992)

A composite source is constituted of the partial sources *X*, *Y*, and *Z*. The joint probabilities $P(x_i,y_j)$ of sources *X* and *Y* are described by the following table,

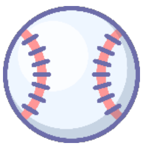
Probabilities		y_i		
		a	***b***	***c***
x_i	***a***	1/16	1/4	1/16
	b	1/16	1/8	1/4
	c	1/16	1/16	1/16

Source *Z*, which is independent of *X* and *Y*, can be represented by the following probability vector,

$$Z = \begin{Bmatrix} a & b & c \\ 1/2, & 1/4, & 1/4 \end{Bmatrix}$$

Find the joint entropy *H(X,Y,Z)* and the conditional entropies *H(X|Y)*, *H(Y|X)*, *H(X|Z)*, *H(Z|X)*, *H(Y|Z)*, and *H(Z|Y)*.

Problem 7 (Cover and Thomas, 2006)



The World Series is a seven-game series that terminates as soon as either team wins four games. Let *X* be the random variable that represents the outcome of a World Series between teams *A* and *B*; possible values of *X* are *AAAA*, *BABABAB*, and *BBBAAAA*. Let *Y* be the number of games played, which ranges from 4 to 7. Assuming that *A* and *B* are equally matched and that the games are independent, calculate *H(X)*, *H(Y)*, *H(Y|X)*, and *H(X|Y)*.

Problem 8 (Lathi and Ding, 2009)

Problem 8.1: A television picture is composed of approximately 300,000 basic picture elements (about 600 picture elements in a horizontal line and 500 horizontal lines per frame). Each of these elements can assume 10 distinguishable brightness levels (such as black and shades of gray) with equal probability. Find the information content of a television picture frame in bits.

Problem 8.2: A radio announcer describes a television picture orally in 1000 words out of his vocabulary of 10,000 words. Assume that each of the 10,000 words in his vocabulary is equally likely to occur in the description of this picture (a crude approximation, but good enough to give an idea). Determine the amount of information broadcast by the announcer in describing the picture. Would you say the announcer can do justice to the picture in 1000 words? Is the old adage "*a picture is worth a thousand words*" an exaggeration or an underrating of the reality?

Problem 9 (Proakis and Salehi, 2008)

Problem 9.1.1: Let *X* be a geometrically distributed random variable, i.e.,

$$P(X=k) = p(1-p)^{k-1} \;\; ; \; k = 1,2,3,\ldots$$

Find the entropy of *X*.

Find the differential entropy of the continuous random variable X in the following cases:

Problem 9.2.1: X is an exponential random variable with parameter $\lambda > 0$, i.e.,

$$p(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Problem 9.2.2: X is a Laplacian random variable with parameter $\lambda > 0$, i.e.,

$$p(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

Problem 9.2.3: X is a triangular random variable with parameter $\lambda > 0$, i.e.,

$$p(x) = \begin{cases} (x+\lambda)/\lambda^2 \ ; & -\lambda \le x \le 0 \\ (-x+\lambda)/\lambda^2 \ ; & 0 < x \le \lambda \\ 0 \ ; & \text{otherwise} \end{cases}$$

Problem 10 (Cover and Thomas, 2006)

Consider the discrete memoryless channel $Y = X + Z$ (mod 11), where

$$Z = \begin{pmatrix} 1 & 2 & 3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

and $X \in \{0,1, ..., 10\}$. Assume that Z is independent of X.

Problem 10.1: Find the capacity.

Problem 10.2: What is the maximizing probability $p^*(x)$?

Problem 11 (Cover and Thomas, 2006)

Problem 11.1: Consider a 26-key typewriter. If pushing a key results in printing the associated letter, what is the capacity C in bits?

Problem 11.2: Now suppose that pushing a key results in printing that letter or the next (with equal probability). Thus $A \to A$ or B, ..., $Z \to Z$ or A. What is the capacity?

Problem 11.3: What is the largest rate code with block length one that you can find that achieves *zero* probability of error for the channel in part 2?

Problem 12 (Cover and Thomas, 2006)

The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \ ; \ x, y \in \{0,1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

Problem 13 (Cover and Thomas, 2006)

Channels with dependence between the letters. Consider the following channel over a binary alphabet that takes in 2-bit symbols and produces a 2-bit output, as determined by the following mapping: 00 → 01, 01 → 10, 10 → 11, and 11 → 00. Thus, if the 2-bit sequence 01 is the input to the channel, the output is 10 with probability 1. Let X_1, X_2 denote the two input symbols and Y_1, Y_2 denote the corresponding output symbols.

Problem 13.1: Calculate the mutual information $I(X_1,X_2; Y_1,Y_2)$ as a function of the input distribution on the four possible pairs of inputs.

Problem 13.2: Show that the capacity of a pair of transmissions on this channel is 2 bits.

Problem 13.3: Show that under the maximizing input distribution, $I(X_1;Y_1) = 0$. Thus, the distribution on the input sequences that achieves capacity does not necessarily maximize the mutual information between individual symbols and their corresponding inputs.

Problem 14 (Proakis and Salehi, 2006)

It can be shown that the rate distortion function for a Laplacian source $p(x) = (2\lambda)^{-1}exp(-|x|/\lambda)$ with an absolute value of error distortion measure $d(x,\hat{x}) = |x - \hat{x}|$ is given by

$$R(D)=\begin{cases}\log_2(\lambda/D) \ ; \ 0\le D\le\lambda \\ 0 \ ; \ D>\lambda\end{cases}$$

Problem 14.1: How many bits per sample are required to represent the outputs of this source with an average distortion not exceeding $\lambda/2$?
Problem 14.2: Plot $R(D)$ for three different values of λ, and discuss the effect of changes in λ on these plots.

Problem 15

In a communication system, the source has an alphabet of five letters x_i, namely $\{x_1, x_2, x_3, x_4, x_5\}$, which are assigned probabilities {0.3, 0.15, 0.25, 0.05, 0.25}. Use the Huffman encoding procedure to determine a binary code for the source output. Find the average code word length, the entropy of the code, and its efficiency.

Problem 16

In a communication system, the source has an alphabet of six letters y_i, namely $\{y_1, y_2, y_3, y_4, y_5, y_6\}$, which are assigned probabilities {0.1, 0.4, 0.06, 0.1, 0.04, 0.3}. Use the Huffman encoding procedure to determine a binary code for the source output. Find the average code word length, the entropy of the code, and its efficiency.

Problem 17

Generate a Shannon-Fano code for a random message with four symbols $\{x_1, x_2, x_3, x_4\}$ which are assigned the probability vector p = {0.5, 0.25, 0.125, 0.125}. Find the average code word length, the entropy of the code, and its efficiency.

Problem 18

Generate a Shannon-Fano code for a random message with five symbols $\{x_1, x_2, x_3, x_4, x_5\}$ which are assigned the probability vector p = {0.35, 0.25, 0.15, 0.15, 0.1}. Find the average code word length, the entropy of the code, and its efficiency.

SOLUTIONS

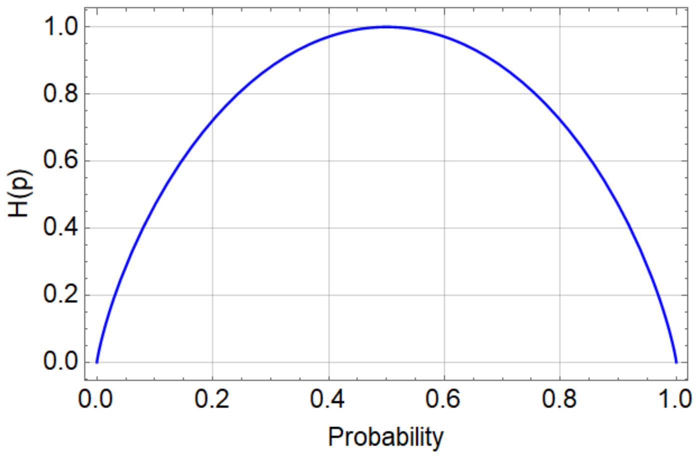
P.1 ➡ *Solution*

1.True. The relation between the two random variables is one-to-one and hence the entropy – which is a function of the probabilities and not the values of a random variable – does not change as one RV is expressed in terms of the other. This suffices for us to state that $H(X) = H(Y)$.

2.False. Actually,

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

which, upon being plotted as a function of p, yields a concave-down curve, as shown. It can be seen that the entropy of X reaches a maximum when the probability p equals 1/2.



3.False. Using Mathematica, we integrate the entropy function from p = 0 to p = 1 to obtain

```
In[22]:= Integrate[-p*Log[2, p] - (1 - p)*Log[2, 1 - p], {p, 0, 1}]
Out[22]= 1/Log[4]
```

That is, the mean entropy is $1/\ln 4 = 0.721$ bits. The inattentive student may have marked this statement as true because integrating the entropy function expressed in terms of natural logarithms does yield 1/2:

```
In[24]:= Integrate[-p*Log[E, p] - (1 - p)*Log[E, 1 - p], {p, 0, 1}]
Out[24]= 1/2
```

However, in this case the entropy is expressed in *nats*, not bits.

4.False. Let

$$A = \sum_{n=2}^{\infty} \left(n \log_2^2 n\right)^{-1}$$

It can be shown that the integer-valued random variable X defined by $Pr(X = n) = (An \log_2^2 n)^{-1}$ for $n = 2, 3, \ldots$, has $H(X) = +\infty$.

5.False. The differential entropy of a Gaussian RV can be determined with the relation

$$H(X) = \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) = 0.5 \times \log\left(2\pi \times 2.72 \times 10\right) = \boxed{3.71\,\text{bits}}$$

6.True. In general, the capacity of a binary symmetric channel with bit-error probability e ($0 \le e < 1/2$) is $\gamma = 1 - h(e)$, where

$$h(e) = e\log_2(1/e) + (1-e)\log_2\left(1/(1-e)\right)$$

In the present case,

$$h(e) = 0.01\log_2(1/0.01) + (1-0.01)\log_2\left(1/(1-0.01)\right) = 0.0808$$

so that $\gamma = 1 - 0.0808 = 0.9192$ bits.

7.True. Indeed, ASCII is a fixed length code in which every code word occupies 8 bits.

8.False. Such a code will be possible if it does not violate the Kraft inequality,

$$\sum_{i=1}^{k} 2^{-\ell_i} \le 1$$

In the case at hand,

$$\sum_{i=1}^{5} 2^{-\ell_i} = 2^{-2} + 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} = \underline{0.9375 < 1}$$

Thus, the code in question is possible.

P.2 ➡ *Solution*

The entropy of X can be expressed as

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - P(x_3)\log_2\left[P(x_3)\right] - P(x_4)\log_2\left[P(x_4)\right]$$

$$\therefore H(X) = 1 - P(x_3)\log_2\left[P(x_3)\right] - P(x_4)\log_2\left[P(x_4)\right]$$

Now, since $P(x_1) + P(x_2) + P(x_3) + P(x_4) = 1$, we see that $P(x_3) + P(x_4) = 1/4$, or $P(x_4) = 1/4 - P(x_3)$. Substituting this latter result into the expression above,

$$H(X) = 1 - P(x_3)\log_2\left[P(x_3)\right] - \left[\frac{1}{4} - P(x_3)\right]\log_2\left[\frac{1}{4} - P(x_3)\right]$$

Differentiating with respect to $P(x_3)$ and setting the result to zero,

$$\frac{dH(X)}{dP(x_3)} = -\log_2\left[P(x_3)\right] + \log_2\left[\frac{1}{4} - P(x_3)\right] = 0$$

The derivative vanishes if $P(x_3) = P(x_4) = 1/8$. These are the probabilities needed for maximum entropy; we leave it to the student to take the second derivative of the expression above and check that $d^2H(X)/dP(x_3)^2 < 0$, which indicates that *H(X)* is indeed a maximum at the probability values in question. The maximum entropy is then

$$\left[H(X)\right]_{\max} = -\frac{1}{2}\log_2\left(\frac{1}{2}\right)-\frac{1}{4}\log_2\left(\frac{1}{4}\right)-\frac{1}{8}\log_2\left(\frac{1}{8}\right)-\left(\frac{1}{8}\right)\log_2\left(\frac{1}{8}\right)=\boxed{\frac{7}{4}}$$

P.3 ➡ Solution

The composite source joint entropy is given by the straightforward summation

$$H(X,Y) = -\sum_{i=1}^{3}\sum_{j=1}^{3}P(x_i,y_j)\log_2\left[P(x_i,y_j)\right]$$

$$H(X,Y) = -\frac{4}{15}\log_2\left(\frac{4}{15}\right)-\frac{1}{15}\log_2\left(\frac{1}{15}\right)-\frac{8}{27}\log_2\left(\frac{8}{27}\right)$$

$$-\frac{8}{27}\log_2\left(\frac{8}{27}\right)-\frac{1}{27}\log_2\left(\frac{1}{27}\right)-\frac{4}{135}\log_2\left(\frac{4}{135}\right)-\frac{1}{135}\log_2\left(\frac{1}{135}\right)=\underline{2.189\text{ bit}}$$

Note that we have ignored the contributions to entropy for *P(a,a)* and *P(b,c)*. Now, to compute *H(X)* and *H(Y)*, first note that

$$P(x=a)=P(x=a,y=a)+P(x=a,y=b)+P(x=a,y=c)$$

$$\therefore P(x=a)=0+\frac{4}{15}+\frac{1}{15}=\frac{1}{3}$$

Likewise, $P(x = b) = 8/27 + 8/27 = 16/27$ and $P(x = c) = 1/27 + 4/135 + 1/135 = 2/27$. Proceeding similarly with the probabilities of *y*, we easily obtain $P(y = a) = 1/3$, $P(y = b) = 16/27$, and $P(y = c) = 2/27$. We proceed to compute the entropy components

$$H(X)=H(Y)=-\frac{1}{3}\log_2\left(\frac{1}{3}\right)-\frac{16}{27}\log_2\left(\frac{16}{27}\right)-\frac{2}{27}\log_2\left(\frac{2}{27}\right)=1.254\text{ bit}$$

Finally,

$$H(X)+H(Y)=1.254+1.254=\underline{2.508\text{ bit}}$$

so we may write

$$H(X,Y)<H(X)+H(Y)$$

The joint entropy is lower than the individual entropies because the sources are not independent of each other.

P.4 ➡ Solution

Determining the entropy of *X* is effortless,

$$H(X)=-0.2\log_2 0.2-0.3\log_2 0.3-0.5\log_2 0.5=\boxed{1.48\text{ bit}}$$

Since the sources are independent, we may write

$$H(X,Y)=H(X)+H(Y)=1.48+1.585=\boxed{3.07\text{ bit}}$$

Equipped with the entropy *H(Y)* only, one generally cannot determine the associated probabilities $p(y_j)$. However, source *Y* happens to have the maximum entropy for a source with $n = 3$ symbols, which is $\log_2 n = \log_2 3 = 1.585$ bit. Since entropy is maximum for equiprobable symbols, we surmise that $P(y = a) = P(y = b) = P(y = c) = 1/3$. Combining these probabilities with the given $P(x_i)$, we can establish the desired joint probabilities.

Probabilities		y_i		
		a	*b*	*c*
x_i	*a*	0.0667	0.0667	0.0667
	b	0.1	0.1	0.1
	c	0.166	0.166	0.166

P.5 ➡ Solution

From the probability vector of *Y*, it is apparent that all columns of the joint probability matrix must add up to 1/4. Using this reasoning, we can easily establish the missing data, as shown on the next page.

Probabilities		y_i			
		a	*b*	*c*	*d*
x_i	*a*	1/8	1/8	1/8	1/8
	b	0	0	0	0
	c	0	1/8	0	1/8
	d	1/8	0	1/8	0

The conditional entropies *H(Y|X)* and *H(X|Y)* can be determined from the general relationships

$$H(Y|X) = H(X,Y) - H(X) \quad \text{(I)}$$

and

$$H(X|Y) = H(X,Y) - H(Y) \quad \text{(II)}$$

From the probability matrix, it is apparent that $P(x = a) = 1/2$, $P(x = b) = 0$, $P(x = c) = 1/4$, and $P(x = d) = 1/4$, giving

$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 1.5 \text{ bit}$$

Referring to the probability vector for *Y*,

$$H(Y) = -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 2.0 \text{ bit}$$

The joint entropy is, in turn,

$$H(X,Y) = -\sum_{i=1}^{4}\sum_{j=1}^{4} P(x_i, y_i)\log_2\left[P(x_i, y_j)\right] = 8 \times \frac{1}{8}\log_2 8 = 3.0 \text{ bit}$$

We substitute in (I) and (II) to obtain

$$H(Y|X) = H(X,Y) - H(X) = 3.0 - 1.5 = \boxed{1.5 \text{ bit}}$$

and

$$H(X|Y) = H(X,Y) - H(Y) = 3.0 - 2.0 = \boxed{1.0 \text{ bit}}$$

P.6 ➡ *Solution*

The probability vectors associated with *X* and *Y* are easily determined to be

$$X = \begin{Bmatrix} a & b & c \\ 6/16, & 7/16, & 3/16 \end{Bmatrix} \; ; \; Y = \begin{Bmatrix} a & b & c \\ 3/16, & 7/16, & 6/16 \end{Bmatrix}$$

so that

$$H(X) = -\frac{6}{16}\log_2\left(\frac{6}{16}\right) - \frac{7}{16}\log_2\left(\frac{7}{16}\right) - \frac{3}{16}\log_2\left(\frac{3}{16}\right) = 1.51 \text{ bit}$$

$$H(Y) = -\frac{3}{16}\log_2\left(\frac{3}{16}\right) - \frac{7}{16}\log_2\left(\frac{7}{16}\right) - \frac{6}{16}\log_2\left(\frac{6}{16}\right) = 1.51 \text{ bit}$$

Also,

$$H(Z) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 1.50 \text{ bit}$$

and

$$H(X,Y) = -\sum_{i=1}^{3}\sum_{j=1}^{3} P(x_i, y_i)\log_2\left[P(x_i, y_j)\right]$$

$$\therefore H(X,Y) = -6 \times \frac{1}{16} \times \log_2\left(\frac{1}{16}\right) - 2 \times \frac{1}{4} \times \log_2\left(\frac{1}{4}\right) - \frac{1}{8} \times \log_2\left(\frac{1}{8}\right) = 2.88 \text{ bit}$$

Since source *Z* is independent of *X* and *Y*, we may write

$$H(X,Y,Z) = H(X,Y) + H(Z) = 2.88 + 1.50 = \boxed{4.38 \text{ bit}}$$

Further, the fact that *Z* is independent of the two other sources enables us to write $H(X,Z) = H(X) + H(Z)$ and $H(Y,Z) = H(Y) + H(Z)$. It follows that

$$H(X|Y) = H(X,Y) - H(Y) = 2.88 - 1.51 = \boxed{1.37 \text{ bit}}$$

$$H(Y|X) = H(X,Y) - H(X) = 2.88 - 1.51 = \boxed{1.37 \text{ bit}}$$

$$H(X|Z) = H(X,Z) - H(Z) = [H(X) + H(Z)] - H(Z) = \boxed{1.51 \text{bit}}$$

$$H(Z|X) = H(X,Z) - H(X) = [H(X) + H(Z)] - H(X) = \boxed{1.50 \text{bit}}$$

$$H(Y|Z) = H(Y,Z) - H(Z) = [H(Y) + H(Z)] - H(Z) = \boxed{1.51 \text{bit}}$$

$$H(Z|Y) = H(Z,Y) - H(Y) = [H(Z) + H(Y)] - H(Y) = \boxed{1.50 \text{bit}}$$

P.7 ➡ *Solution*

There are 2 World Series with 4 games, namely *AAAA* and *BBBB*; each happens with probability $(1/2)^4$. There are 8 (= $2C_4^3$) World Series with 5 games; each happens with probability $(1/2)^5$. There are 20 (= $2C_5^3$) World Series with 6 games; each happens with probability $(1/2)^6$. There are 40 (= $2C_6^3$) World Series with 7 games; each happens with probability $(1/2)^7$.

It follows that the probability of a 4-game series ($Y = 4$) is $2 \times (1/2)^4 = 1/8$; the probability of a 5-game series ($Y = 5$) is $8 \times (1/2)^5 = 1/4$; the probability of a 6-game series ($Y = 6$) is $20 \times (1/2)^6 = 5/16$; the probability of a 7-game series ($Y = 7$) is $40 \times (1/2)^7 = 5/16$. The entropy of X, the variable that represents the outcome of a World Series between teams *A* and *B*, is

$$H(X) = \sum p(x) \log_2 \left[\frac{1}{p(x)}\right]$$

$$\therefore H(X) = 2\left(\frac{1}{2}\right)^4 \log_2(2^4) + 8\left(\frac{1}{2}\right)^5 \log_2(2^5) + 20\left(\frac{1}{2}\right)^6 \log_2(2^6) + 40\left(\frac{1}{2}\right)^7 \log_2(2^7)$$

$$\therefore \boxed{H(X) = 5.81 \text{ bit}}$$

The entropy of Y, the variable that represents the number of games played, is

$$H(Y) = \sum p(y) \log_2 \left[\frac{1}{p(y)}\right]$$

$$\therefore H(Y) = \left(\frac{1}{8}\right)\log_2(8) + \left(\frac{1}{4}\right)\log_2(4) + \left(\frac{5}{16}\right)\log_2\left(\frac{16}{5}\right) + \left(\frac{5}{16}\right)\log_2\left(\frac{16}{5}\right) = \boxed{1.92 \text{bit}}$$

Y is a deterministic function of X, which is to say that if we know X, there is no randomness in Y. Thus, conditional entropy $H(Y|X) = 0$. Now, if

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

we may write

$$H(X|Y) = H(X) + H(Y|X) - H(Y) = 5.81 + 0 - 1.92 = \boxed{3.89 \text{bit}}$$

P.8 ➡ *Solution*

Problem 8.1: The amount of information in one picture element is $\log_2 10 = 3.32$ bits. The amount of information in a picture frame follows as 300,000 × 3.32 = 996,000 bits.

Problem 8.2: Given the announcer's vocabulary of 10,000 words, the information content of a word is estimated at $\log_2 10{,}000 = 13.3$ bits. It follows that a thousand words amount to 1000 × 13.3 = 13,300 bits. We estimated above that the information per picture frame stands at nearly 1 million bits, which is about two orders of magnitude greater than the information content estimated for 1000 words. Hence, to say that a picture is worth 1000 words is very much an underrating of the real situation.

P.9 ➡ *Solution*

Problem 9.1.1: The entropy of a discrete random variable is given by

$$H(X) = -\sum p(x) \log p(x)$$

In the present case,

$$H(X) = -\sum_{k=1}^{\infty} p(1-p)^{k-1} \log_2\left[p(1-p)^{k-1}\right]$$

Expanding the logarithm of a product into a sum of logarithms and manipulating,

$$H(X) = -\sum_{k=1}^{\infty} p(1-p)^{k-1}\left[\log_2 p + \log_2 (1-p)^{k-1}\right]$$

$$\therefore H(X) = -p\log_2 p\sum_{k=1}^{\infty}(1-p)^{k-1} - p\log_2(1-p)\sum_{k=1}^{\infty}(k-1)(1-p)^{k-1}$$

The first summation on the right-hand side is the summation of an infinite geometric series with first term 1 and common ratio (1 – p). The second summation can be established with the Mathematica code

```
In[62]:= Simplify[Sum[(k - 1) * (1 - p)^(k-1), {k, 1, ∞}]]
Out[62]= (1 - p)/p^2
```

Accordingly,

$$H(X) = -p\log_2 p \times \frac{1}{1-(1-p)} - p\log_2(1-p)\times\frac{1-p}{p^2}$$

$$\therefore \boxed{H(X) = -\log_2 p - \frac{(1-p)}{p}\log_2(1-p)}$$

Problem 9.2.1: In general, the differential entropy can be defined as

$$H(x) = -\int p(x)\ln\left[p(x)\right]dx$$

so that, for an exponential random variable,

$$H(x) = -\int_0^{\infty}\frac{1}{\lambda}e^{-x/\lambda}\ln\left(\frac{1}{\lambda}e^{-x/\lambda}\right)dx$$

$$\therefore H(x) = -\int_0^{\infty}\frac{1}{\lambda}e^{-x/\lambda}\left[\ln\left(\frac{1}{\lambda}\right) + \underbrace{\ln e^{-x/\lambda}}_{-x/\lambda}\right]dx$$

$$\therefore H(x) = -\ln\left(\frac{1}{\lambda}\right)\int_0^{\infty}\frac{1}{\lambda}e^{-x/\lambda}dx + \frac{1}{\lambda}\int_0^{\infty}e^{-x/\lambda}\left(\frac{x}{\lambda}\right)dx$$

The first integral on the right-hand side is elementary and equals 1; the second integral is actually the first moment or expected value of the exponential random variable and hence equals the parameter λ, so that

$$H(x) = -\ln\left(\frac{1}{\lambda}\right)\times 1 + \frac{1}{\lambda}\times\lambda = \boxed{\ln(\lambda)+1}$$

Problem 9.2.2: From the definition of differential entropy, we manipulate to obtain

$$H(x) = -\int_{-\infty}^{\infty}\frac{1}{2\lambda}e^{-|x|/\lambda}\ln\left(\frac{1}{2\lambda}e^{-|x|/\lambda}\right)dx$$

$$\therefore H(x) = -\int_{-\infty}^{\infty}\frac{1}{2\lambda}e^{-|x|/\lambda}\left[\ln\left(\frac{1}{2\lambda}\right) + \underbrace{\ln e^{-|x|/\lambda}}_{=-|x|/\lambda}\right]dx$$

$$\therefore H(x) = -\ln\left(\frac{1}{2\lambda}\right)\underbrace{\int_{-\infty}^{\infty}\frac{1}{2\lambda}e^{-|x|/\lambda}dx}_{=1} + \frac{1}{2\lambda}\int_{-\infty}^{\infty}\frac{|x|}{\lambda}e^{-|x|/\lambda}dx$$

$$\therefore H(x) = \ln(2\lambda) + \frac{1}{2\lambda^2}\int_{-\infty}^{\infty}|x|e^{-|x|/\lambda}dx$$

Using the definition of absolute value, we restate the remaining integral and evaluate, giving

$$H(x) = \ln(2\lambda) + \frac{1}{2\lambda^2}\left[\int_{-\infty}^{0}-xe^{x/\lambda}dx + \int_0^{\infty}xe^{-x/\lambda}dx\right]$$

$$\therefore H(x) = \ln(2\lambda) + \frac{1}{2\lambda^2}\left(\lambda^2+\lambda^2\right) = \boxed{\ln(2\lambda)+1}$$

Problem 9.2.3: In this case, the differential entropy is determined as

$$H(X)=-\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln\left(\frac{x+\lambda}{\lambda^2}\right)dx-\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln\left(\frac{-x+\lambda}{\lambda^2}\right)dx$$

$$\therefore H(X)=-\ln\left(\frac{1}{\lambda^2}\right)\left(\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}dx+\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}dx\right)$$

$$-\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln(x+\lambda)dx-\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln(-x+\lambda)dx$$

$$\therefore H(X)=-\ln\left(\frac{1}{\lambda^2}\right)\left(\frac{1}{2}+\frac{1}{2}\right)-\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln(x+\lambda)dx-\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln(-x+\lambda)dx$$

$$\therefore H(X)=-\ln\left(\frac{1}{\lambda^2}\right)-\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln(x+\lambda)dx-\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln(-x+\lambda)dx$$

The remaining integrals yield the same results, namely

$$\int_{-\lambda}^{0}\frac{x+\lambda}{\lambda^2}\ln(x+\lambda)dx=\frac{1}{4}\left[2\ln(\lambda)-1\right]$$

$$\int_{0}^{\lambda}\frac{-x+\lambda}{\lambda^2}\ln(-x+\lambda)dx=\frac{1}{4}\left[2\ln(\lambda)-1\right]$$

so that

$$H(X)=-\ln\left(\frac{1}{\lambda^2}\right)-\frac{1}{4}\left[2\ln(\lambda)-1+2\ln(\lambda)-1\right]$$

$$\therefore \boxed{H(X)=\ln(\lambda^2)-\ln(\lambda)+\frac{1}{2}}$$

P.10 ➡ *Solution*

Problem 10.1: We have $Y = X + Z(\text{mod } 11)$ and

$$Z=\begin{cases}1,\text{ with probability } 1/3\\2,\text{ with probability } 1/3\\3,\text{ with probability } 1/3\end{cases}$$

We then have the entropies

$$H(Y\,|\,X)=H(Z\,|\,X)=H(Z)=\log_2 3$$

that is, the values are independent of the distribution of *X*, and hence the capacity of the channel is

$$C=\max_{p(x)} I(X;Y)=\max_{p(x)}\left[H(Y)-H(Y\,|\,X)\right]$$

$$\therefore \boxed{C=\log_2 11-\log_2 3}$$

which is attained when *Y* has a uniform distribution; by symmetry, this occurs when *X* has a uniform distribution. The capacity of the channel is $\log_2(11/3)$ bits/transmission.

Problem 10.2: The capacity in question is achieved when the inputs have an uniform distribution, i.e., $p(X = i) = 1/11$, for $i = 0,1, ..., 10$.

P.11 ➡ *Solution*

Problem 11.1: If the typewriter prints whatever key is struck, then the output, *Y*, is the same as the input, *X*, and

$$C=\max I(X;Y)=\max H(X)=\boxed{\log_2 26}$$

This is the capacity, which is attained by a uniform distribution over the letters.

Problem 11.2: In this case, the output is either equal to the input (with probability 1/2) or equal to the next letter (with probability 1/2). Hence $H(Y|X) = \log_2 2$ is independent of the distribution of *X*, and hence

$$C=\max I(X;Y)=\max\left[H(Y)-H(Y\,|\,X)\right]=\log_2 26-\log_2 2=\boxed{\log_2 13}$$

attained for a uniform distribution over the output, which in turn is attained by a uniform distribution on the input.

Problem 11.3: A simple zero error block length one code is the one that uses every alternate letter, as in *A, C, E, G, ..., W, Y*. In this case, none of the code words will be confused, since A will produce either A or B, C will produce either C or D, etc. The rate of this code is

$$R = \frac{\log_2(\text{No. of code words})}{\text{Block length}} = \frac{\log_2 13}{1} = \boxed{\log_2 13}$$

In this case, we can achieve capacity using a simple code with zero error.

P.12 ➡ Solution

First we express $I(X;Y)$, the mutual information between the input and output of the Z-channel, as a function of $x = \Pr(X = 1)$,

$$H(Y|X) = \Pr(X=0)\times 0 + \Pr(X=1)\times 1 = x$$
$$H(Y) = H(\Pr(Y=1)) = H(x/2)$$
$$I(X;Y) = H(Y) - H(Y|X) = H(x/2) - x$$

Since $I(X;Y) = 0$ when $x = 0$ and $x = 1$, the maximum mutual information is obtained for some value of x such that $0 < x < 1$. Using elementary calculus, we can establish

$$\frac{d}{dx} I(X;Y) = \frac{1}{2}\log_2\left(\frac{1-x/2}{x/2}\right) - 1$$

which is equal to zero for $x = 2/5$. (It is reasonable that $Pr(X = 1) < 1/2$ because $X = 1$ is the noisy input to the channel.) It follows that the capacity of the Z-channel is

$$I(X;Y) = H(0.4/2) - 0.4 = 0.722 - 0.4 = \boxed{0.322}$$

P.13 ➡ Solution

Problem 13.1: If we look at pairs of inputs and pairs of outputs, this channel is a noiseless four input-four output channel. Let the probabilities of the four input pairs be p_{00}, p_{01}, p_{10}, p_{11}, respectively; the mutual information is then

$$I(X_1,X_2;Y_1,Y_2) = H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2)$$
$$\therefore I(X_1,X_2;Y_1,Y_2) = H(Y_1,Y_2) - 0$$
$$\therefore \boxed{I(X_1,X_2;Y_1,Y_2) = H(p_{11},p_{00},p_{01},p_{10})}$$

Problem 13.2: The capacity of the channel is achieved by a uniform distribution over the inputs, which produces a uniform distribution on the output pairs, so we may write

$$C = \max_{p(x_1,x_2)} I(X_1,X_2;Y_1,Y_2) = 2\,\text{bits}$$

and the maximizing $p(x_1,x_2)$ puts probability 1/4 on each of the pairs 00, 01, 10, and 11.

Problem 13.3: To calculate the mutual information $I(X_1,Y_1)$, we need to calculate the joint distribution of X_1 and Y_1. The joint distribution of X_1X_2 and Y_1Y_2 under an uniform input distribution is given by the following matrix.

X_1X_2/Y_1Y_2	00	01	10	11
00	0	1/4	0	0
01	0	0	1/4	0
10	0	0	0	1/4
11	1/4	0	0	0

From this table, it is easy to establish the joint distribution of X_1 and Y_1, as shown.

X_1/Y_1	0	1
0	1/4	1/4
1	1/4	1/4

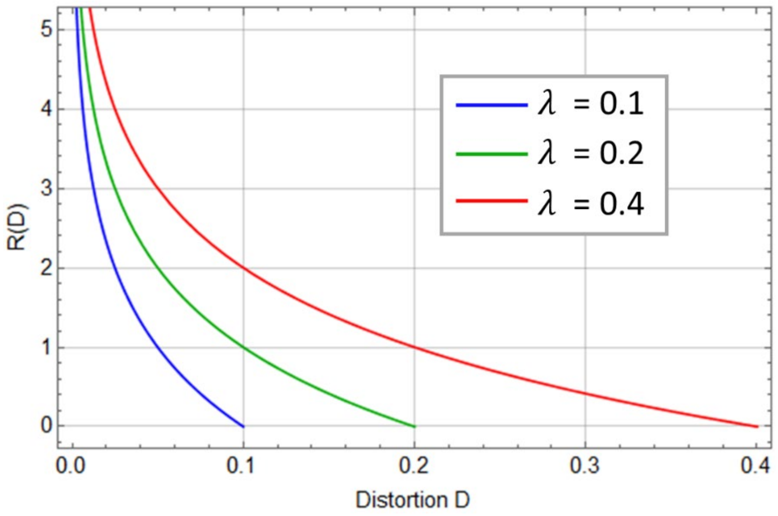
Clearly, the marginal distributions of X_1 and Y_1 are both (1/2, 1/2) and, in view of the fact that the joint distribution is the product of the marginals, X_1 is independent of Y_1 and $I(X_1;Y_1) = 0$. Thus the distribution on the input sequences that achieves capacity does not necessarily maximize the mutual information between individual symbols and their corresponding outputs.

P.14 ➡ *Solution*

Problem 14.1: Since *R*(*D*) = *log*(*λ*/*D*) and *D* = *λ*/2, we obtain

$$R(D) = \log_2\left(\frac{\lambda}{\lambda/2}\right) = \boxed{1\,\text{bit/sample}}$$

Problem 14.2: *R*(*D*) is plotted for three values of *λ* in the following figure. As can be seen, an increase in the parameter *λ* increases the rate for a given distortion.



P.15 ➡ *Solution*

Let us work out the procedure step by step. We first order the letters vertically in descending order of probability.

x_1 0.3

x_3 0.25

x_5 0.25

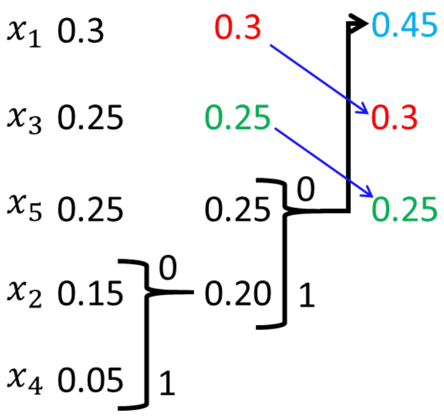
x_2 0.15

x_4 0.05

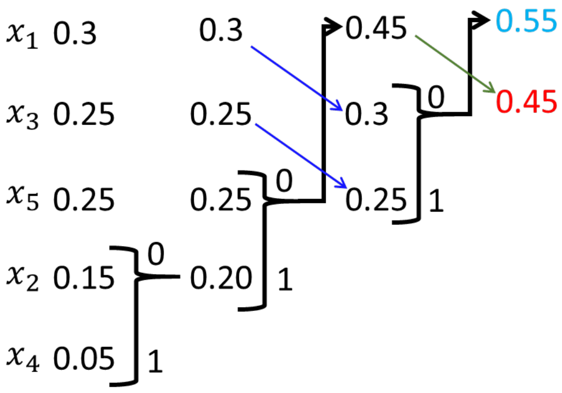
Then, we group the two lowermost symbols and assign 0 to the greater one and 1 to the lower one, as shown. The probabilities of the two least probable symbols are added to form a single symbol, which is then carried over to the second column in a manner that conserves the decreasing order of probabilities. In the present case, this implies that the added probability 0.15 + 0.05 = 0.20 should occupy the bottommost position, because it is less than 0.25.

x_1	0.3		0.3
x_3	0.25		0.25
x_5	0.25		0.25
x_2	0.15	0	0.20
x_4	0.05	1	

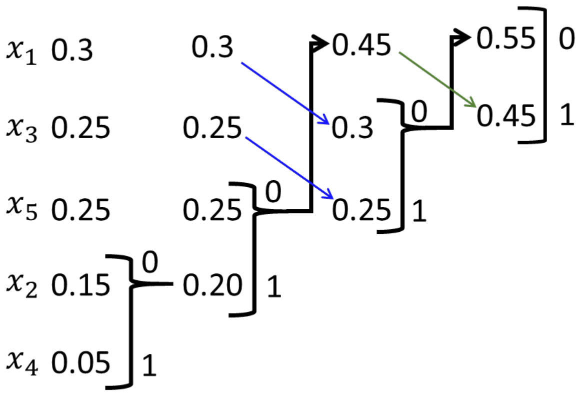
Next, we proceed as we did above and assign a 0 and a 1 to the lowermost probabilities. Then, we add the two lowermost probabilities to obtain 0.25 + 0.20 = 0.45. This new probability is carried over to the third column in a manner that conserves the decreasing order of probabilities. Since 0.45 > 0.3 > 0.25, this new probability should occupy the topmost position in the third column while the remaining probabilities should be displaced downwards, as shown by the dark blue arrows.



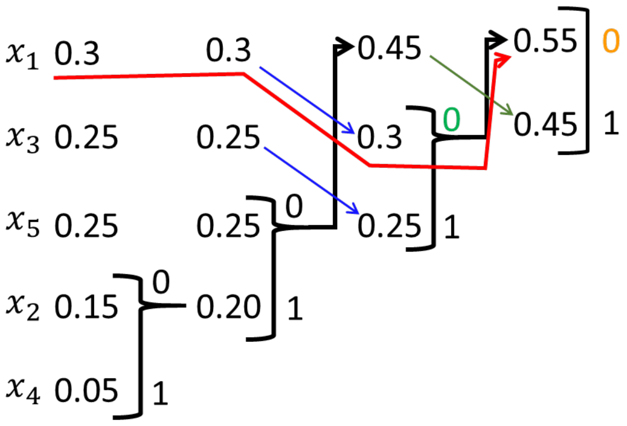
Next, we assign 0 and 1 to the lowermost symbols. Adding the two lowermost probabilities, we get 0.3 + 0.25 = 0.55. Since 0.55 > 0.45, this new probability takes the top spot in the fourth column and the 0.45-probability is displaced downward, as shown by the dark green arrow.



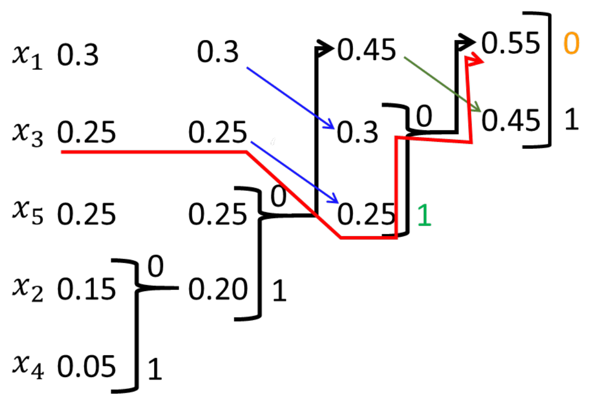
Lastly, a 0 and 1 are assigned to the two remaining symbols, as shown.



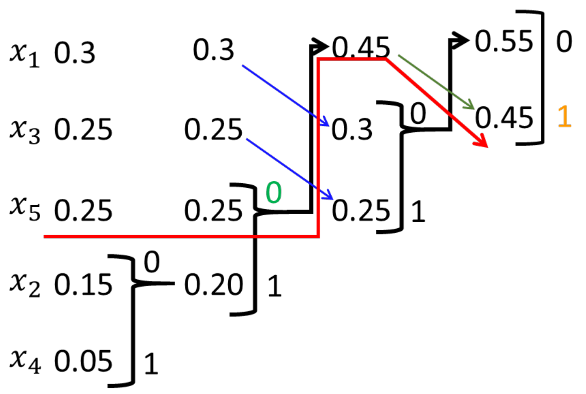
To obtain the code words for each symbol, we trace the probabilities from the leftmost column to the rightmost column, all the while following the pertaining arrows. The code word is obtained by reading, from right to left, the numbers we found in the traced path. For example, in the case of x_1, whose path is shown in red, we pass through a 0 in the fourth column and a 0 in the third column; accordingly, the code word for x_1 is 00.



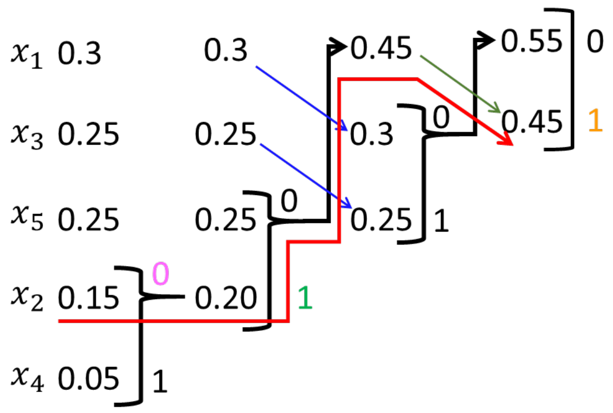
Proceeding similarly with x_3, the traced path, also shown in red, passes through a 0 in the fourth column and a 1 in the third column; it follows that the code word for x_3 is 01.



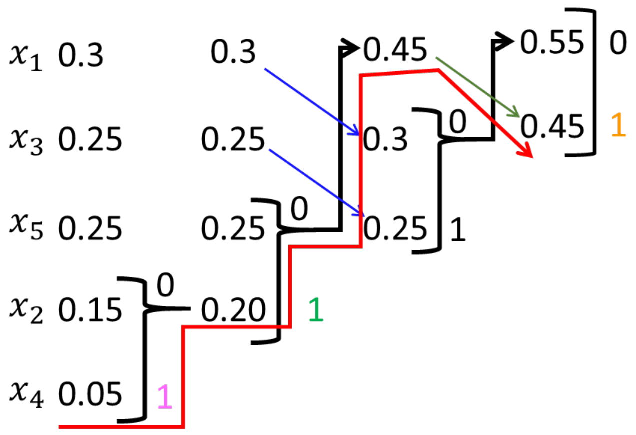
Proceeding in the same manner with x_5, we read a code word 10, as shown.



Proceeding in the same manner with x_2, we read a code word 110, as shown.



Finally, we outline the path for symbol x_4 to read the code word 111.



The code words, probabilities and lengths are summarized below.

Symbol	Prob.	Code	Length
x_1	0.3	00	2
x_2	0.15	110	3
x_3	0.25	01	2
x_4	0.05	111	3
x_5	0.25	10	2

The average code word length is

$$\overline{L} = 2\times 0.3 + 3\times 0.15 + 2\times 0.25 + 3\times 0.05 + 2\times 0.25 = \boxed{2.2\,\text{letters/message}}$$

The entropy is

$$H = -\left(0.3\log_2 0.3 + 0.15\log_2 0.15 + 0.25\log_2 0.25 + 0.05\log_2 0.05 + 0.25\log_2 0.25\right)$$

$$\therefore \boxed{H = 2.148\,\text{bits/symbol}}$$

Finally, the efficiency is expressed as

$$\eta = \frac{H}{\bar{L}\log_2 M} = \frac{2.148}{2.2\log_2 2} = \boxed{97.6\%}$$

P.16 ➡ *Solution*

In the same vein of the previous problem, we'll work out the Huffman encoding procedure step by step. As before, we first order the letters vertically in descending order of probability.

y_2 0.4

y_6 0.3

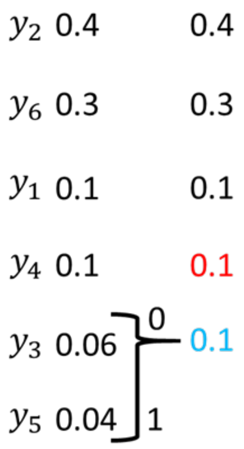
y_1 0.1

y_4 0.1

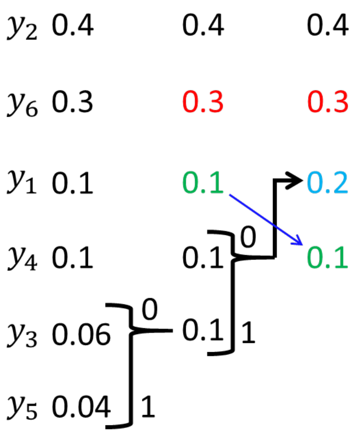
y_3 0.06

y_5 0.04

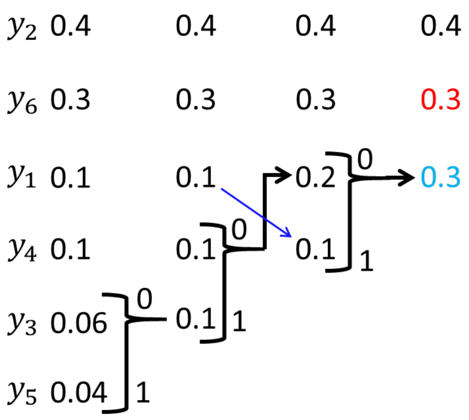
Then, we join the two lowermost symbols, assigning 0 to the greater one and 1 to the lower one, as shown. The probabilities of the two least probable symbols are added to form a single symbol, which is then carried over to the second column in a manner that conserves the decreasing order of probabilities. In the operation at hand, we perform the sum 0.06 + 0.04 = 0.10. Note that there are already two 0.10's in the probability vector; in such a case, we must place the 0.10 obtained via the addition operation in the bottommost position relatively to other data. Pay attention to this rule, because there are numerous solved examples out there, including on Youtube, that involve placing similar probabilities on the *topmost* possible position, which is false and ruins the entire encoding procedure.



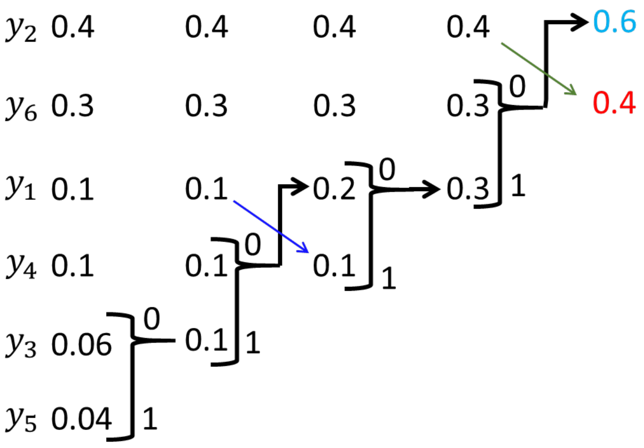
Next, we proceed as we did above and assign a 0 and a 1 to the lowermost probabilities. Then, we add the two lowermost probabilities to obtain 0.10 + 0.10 = 0.20. This new probability is carried over to the third column in a manner that conserves the decreasing order of probabilities. Since 0.30 > 0.20 > 0.10, this new probability should be placed between the 0.30 and the 0.10, as shown.



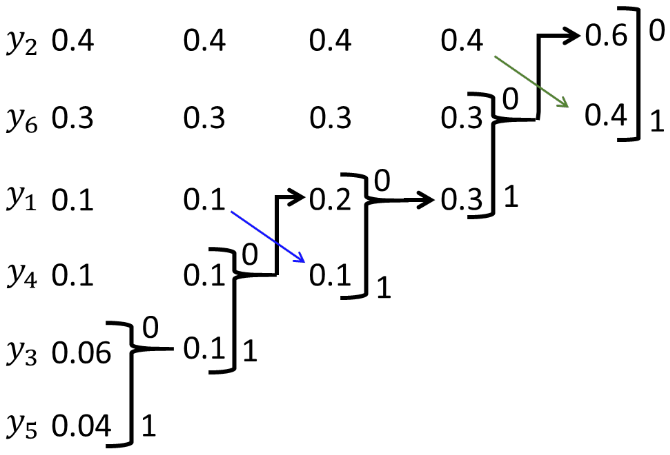
A 0 and a 1 are assigned as usual. In the next calculation, we add 0.2 + 0.1 = 0.30. Note that there is already a 0.3 in the dataset, so we place the 0.30 obtained here in the bottommost possible position, as shown.



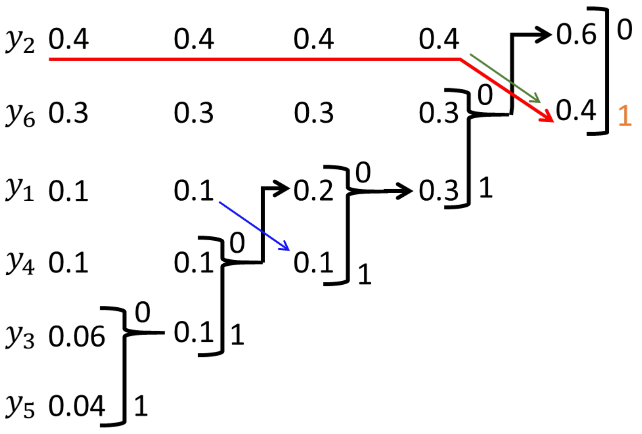
A 0 and a 1 are assigned to the two lowermost symbols. In the next calculation, we add 0.3 + 0.3 = 0.60 and position this result at the top of the next column because 0.60 > 0.40.



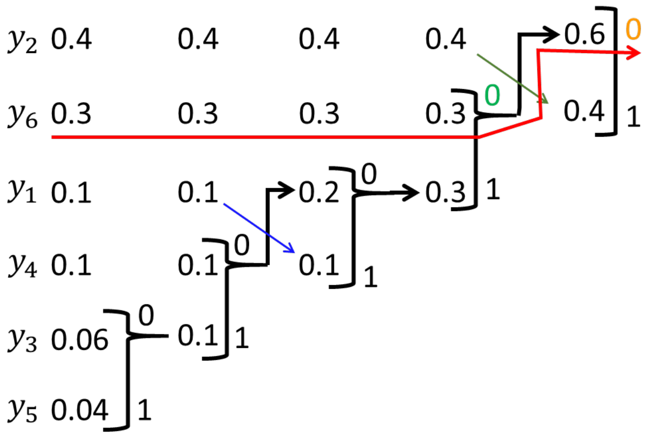
Finally, a 0 and 1 are assigned to the two remaining symbols, as shown.



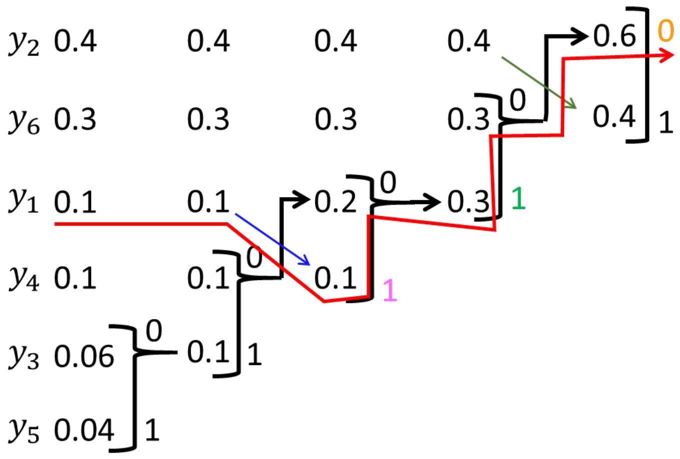
To obtain the code words for each symbol, we trace the probabilities from the leftmost column to the rightmost column, all the while following the pertaining arrows. The code word is obtained by reading, from right to left, the numbers we found in the traced path. For example, in the case of y_2, whose path is shown in red, we pass through a 1 in the fourth column and no other numbers in the remaining columns, as shown. It follows that the code word for y_2 is 1.



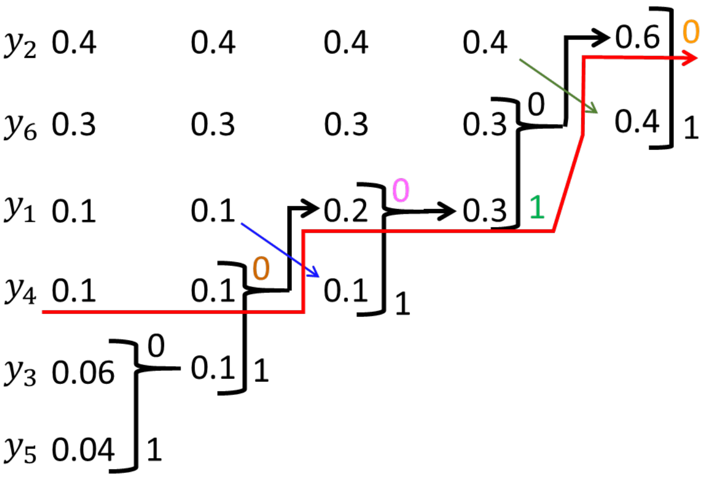
Proceeding similarly with y_6, the traced path, also shown in red, passes through a 0 in the fifth column and a 0 in the fourth column; it follows that the code word for y_6 is 00.



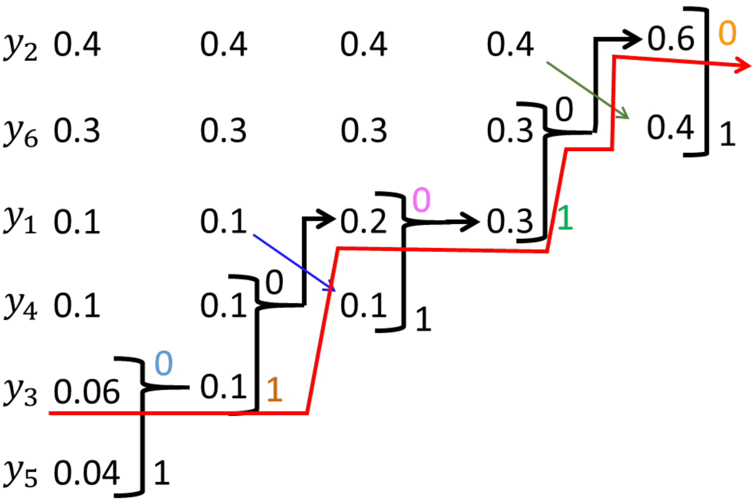
Tracing a path for y_1, we read a code word 011, as shown.



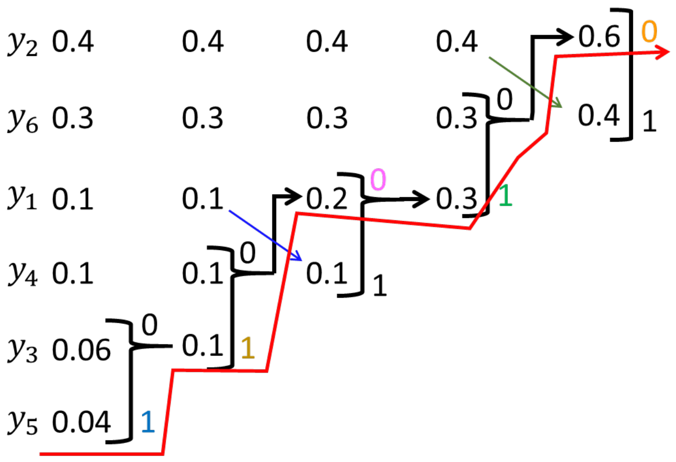
Tracing a path for y_4, we read a code 0100, as shown.



Tracing a path for y_3, we read a code 01010, as shown.



Finally, we trace a path for y_5 to read 01011, as shown.



The code words, probabilities and lengths are summarized below.

Symbol	Prob.	Code	Length
y_1	0.1	011	3
y_2	0.4	00	2
y_3	0.06	01010	5
y_4	0.1	0100	4
y_5	0.04	01011	5
y_6	0.3	1	1

The average code word length is

$$\bar{L} = 3\times 0.1 + 2\times 0.4 + 5\times 0.06 + 4\times 0.1 + 5\times 0.04 + 1\times 0.3 = \boxed{2.3\,\text{letters/message}}$$

The entropy is

$$H = -\begin{pmatrix} 0.1\log_2 0.1 + 0.4\log_2 0.4 + 0.06\log_2 0.06 \\ +0.1\log_2 0.1 + 0.04\log_2 0.04 + 0.3\log_2 0.3 \end{pmatrix} = \boxed{2.144\,\text{bits/symbol}}$$

Finally, the efficiency is expressed as

$$\eta = \frac{H}{\bar{L}\log_2 M} = \frac{2.144}{2.3\log_2 2} = \boxed{93.2\%}$$

P.17 ➡ *Solution*

Stated briefly, the process to determine the Shannon-Fano code for a message involves the following steps:

Step 1: Arrange the symbols in order of nonincreasing probability.

Step 2: Divide the list of ordered symbols into two parts, with the total probability of the left part being as close to the total probability of the right part as possible. In other words, arrange the symbols into two groups that are as equiprobable as possible.

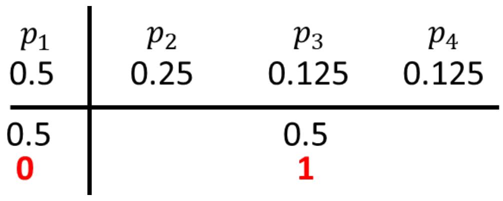
Step 3: Assign the binary digit 0 to the left part of the list, and the digit 1 to the right part. This means that the code words for the symbols in the first part will all start with 0, and the code words for the symbols in the second part will all start with 1.

Step 4: Recursively apply steps 2 and 3 to each of the two parts, subdividing into further parts and assigning bits to the code words until each symbol is the single member of a part.

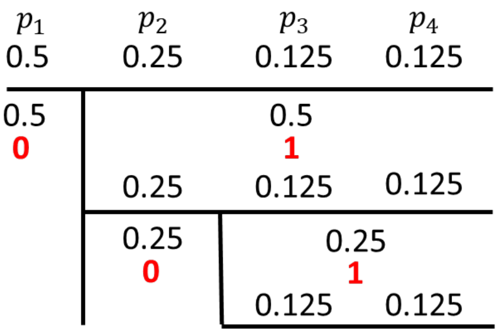
In the present case, the symbols are already organized in order of nonincreasing probability as they are, so we can skip step 1. The symbols are listed below.

p_1	p_2	p_3	p_4
0.5	0.25	0.125	0.125

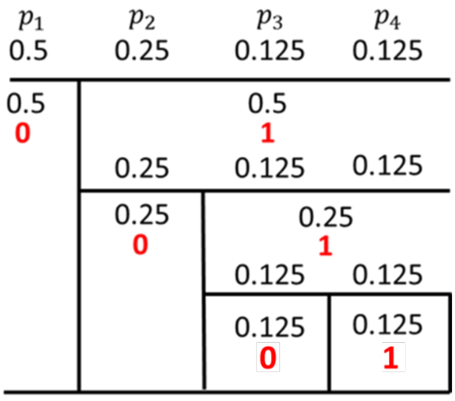
The next step is to divide the symbols into two groups that are as equiprobable as possible. We could group x_1 and x_2 in one set, $\{x_1, x_2\}$, and the other two symbols into another, $\{x_3, x_4\}$. Let us call this the *first possibility*. In this case, the probabilities of the $\{x_1, x_2\}$ set add to 0.5 + 0.25 = 0.75, while the probabilities of the $\{x_3, x_4\}$ set amount to 0.125 + 0.125 = 0.25. Now, another approach is to take symbol x_1 on its own, forming a set $\{x_1\}$, and assign the other three symbols to a second group, $\{x_2, x_3, x_4\}$. Let us call this the *second possibility*. In this case, the probability of the $\{x_1\}$ set is obviously 0.5, while the probabilities of the $\{x_2, x_3, x_4\}$ set add up to 0.25 + 0.125 + 0.125 = 0.5. The two sets thus formed not only have very close probabilities, they are exactly equiprobable. A *third possibility* would be to separate x_4 into a set $\{x_4\}$ and the other three symbols into a set $\{x_1, x_2, x_3\}$, but it is easy to see that this yields two sets with very different total probabilities, which is the opposite of what we're looking for. The second possibility works best. Following step 3, we assign a binary digit 0 to the left part of the list and a digit 1 to the right part, as shown.



Now, the set to the left already has a single entry and cannot be reduced any further. We are left with the set to the right, which has three elements, $\{p_2, p_3, p_4\}$ = {0.25, 0.125, 0.125}. It doesn't take a rocket scientist to see that the most equiprobable two groups of symbols will be obtained if we take x_2 in one set and $\{x_3, x_4\}$ in another, leading to a total probability of 0.25 in the former and a total probability of 0.125 + 0.125 = 0.25 in the latter. We assign a 0 to the set on the left and a 1 to the set on the right, as shown.



We now have another set with a single entry, $\{x_2\}$, so there's no need to work on it any further. On the other hand, there's a set $\{x_3, x_4\}$ that can be subdivided further. Obviously, the one way to subdivide this set is to place the two elements in individual sets of their own, $\{x_3\}$ and $\{x_4\}$, as shown. A 0 is assigned to the set on the left and a 1 is assigned to the set on the right.



The code words are obtained by reading the 0's and 1's from top to bottom. In the case of x_1, for example, there's a single 0 in the entire column, so the code word for this symbol is 0. In the case of x_2, there's a 1 and a 0 as we move from the top of the column to the bottom, so the code word for this symbol is 10. In the case of x_3, in moving from top to bottom we pass through a 1, another 1, and a 0, so the code word for x_3 is 110. Lastly, the code word for x_4 is 111. The results are tabulated below.

Symbol	Prob.	Code	Length
x_1	0.5	0	1
x_2	0.25	10	2
x_3	0.125	110	3
x_4	0.125	111	3

The average code word length is

$$\overline{L} = 1\times 0.5 + 2\times 0.25 + 3\times 0.125 + 3\times 0.125 = \boxed{1.75\,\text{bits/symbol}}$$

The entropy is

$$H = -\left(0.5\log_2 0.5 + 0.25\log_2 0.25 + 0.125\log_2 0.125 + 0.125\log_2 0.125\right) = \boxed{1.75\,\text{bits/symbol}}$$

And the efficiency can only be

$$\eta = \frac{H}{\overline{L}} = \frac{1.75}{1.75} = \boxed{100\%}$$

P.18 ➡ *Solution*

The symbols are already ordered in decreasing order of probability, so there is no need to reorganize them.

p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1

Next, we must split the symbols into two groups that are as equiprobable as possible. One option would be to group symbol x_1 into one set and symbols $\{x_2, x_3, x_4, x_5\}$ into another. In this *first possibility*, the total probability of the set on the left would be 0.35, while that of the set on the right would be 0.25 + 0.15 + 0.15 + 0.1 = 0.65. Another approach would be to group symbols x_1 and x_2 into one set, $\{x_1, x_2\}$, and symbols x_3, x_4, and x_5 into another, $\{x_3, x_4, x_5\}$. In this *second possibility*, the total probability of the set on the left would be 0.35 + 0.25 = 0.60, while that of the set on the right would be 0.15 + 0.15 + 0.1 = 0.40. Yet another approach would be to group symbols x_1, x_2, and x_3 into one set $\{x_1, x_2, x_3\}$ and symbols x_4 and x_5 into another, $\{x_4, x_5\}$. In this *third possibility*, the total probability of the set on the left would be 0.35 + 0.25 + 0.15 = 0.75, while that of the set on the right would be 0.15 + 0.1 = 0.25. A fourth approach would be to form sets $\{x_1, x_2, x_3, x_4\}$ and $\{x_5\}$, but it is apparent that the total probabilities of the two sets thus formed would be substantially different, which is the opposite of what we're looking for. Gleaning our results, it is apparent that the best split is the second possibility, which involves forming two sets $\{x_1, x_2\}$ and $\{x_3, x_4, x_5\}$, as shown. As usual, we assign a 0 to the set on the left and a 1 to the set on the right.

p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1
0.60 0			0.40 1	

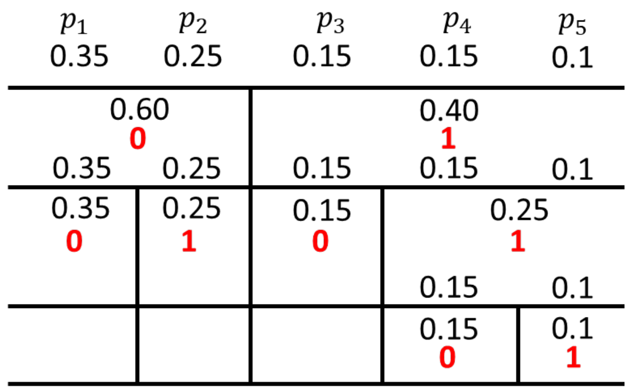
On the left, we now have two probabilities $\{p_1, p_2\} = \{0.35, 0.25\}$. One obvious way to proceed here is to subdivide it into two sets and assign a 0 and a 1 accordingly, as shown.

p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1
0.60 0			0.40 1	
0.35	0.25	0.15	0.15	0.1
0.35 0	0.25 1			

On the right, we have three probabilities $\{p_3, p_4, p_5\} = \{0.15, 0.15, 0.1\}$. To further subdivide this group, we can form a set with x_1 only and another with x_4 and x_5. In this *first possibility*, the total probability of the first set would be 0.15, while that of the second set would be 0.15 + 0.1 = 0.25. A second possible split it to form a set with x_3 and x_4 and another with x_5. In this *second possibility*, the total probability of the first set would be 0.15 + 0.15 = 0.30, while that of the second set would be 0.1. The first possibility leads to the most equiprobable sets, so we form sets $\{x_3\}$ and $\{x_4, x_5\}$ accordingly.

p_1	p_2	p_3	p_4	p_5
0.35	0.25	0.15	0.15	0.1
0.60 0			0.40 1	
0.35	0.25	0.15	0.15	0.1
0.35 0	0.25 1	0.15 0	0.25 1	

At this point, the only group that hasn't been split into a single-entry set is $\{x_4, x_5\}$. As usual, we subdivide this group and ascribe a 0 and a 1 in accordance with step 3 of the algorithm.



The code words are obtained by reading the 0's and 1's from top to bottom. The results are summarized below.

Symbol	Prob.	Code	Length
x_1	0.35	00	2
x_2	0.25	01	2
x_3	0.15	10	2
x_4	0.15	110	3
x_5	0.1	111	3

The average code word length is

$$\bar{L} = 2\times 0.35 + 2\times 0.25 + 2\times 0.15 + 3\times 0.15 + 3\times 0.1 = \boxed{2.25\,\text{bits/symbol}}$$

The entropy is

$$H = -\begin{pmatrix} 0.35\log_2 0.35 + 0.25\log_2 0.25 + 0.15\log_2 0.15 \\ +0.15\log_2 0.15 + 0.1\log_2 0.1 \end{pmatrix} = \boxed{2.183\,\text{bits/symbol}}$$

Lastly, the efficiency of this Shannon-Fano code is

$$\eta = \frac{H}{\bar{L}} = \frac{2.183}{2.25} = \boxed{97.02\%}$$

REFERENCES

- COVER, T.M. and THOMAS, J.A. (2006). *Elements of Information Theory*. 2nd edition. Hoboken: John Wiley and Sons.
- LATHI, B.P. and DING, Z. (2009). *Modern Digital and Analog Communication Systems*. 4th edition. Oxford: Oxford University Press.
- MILDENBERGER, O. (1992). *Informationstheorie und Codierung*. Berlin/Heidelberg: Springer.
- PROAKIS, J.G. and SALEHI, M. (2008). *Digital Communications*. 5th edition. New York: McGraw-Hill.