to 0.5–0.75 bits/pixel results in good to very-good quality images that are sufficient for many applications. At 0.75–1.5 bits/pixel, excellent quality images are obtained sufficient for most applications. Finally, at rates of 1.5–2 bits/pixel, the resulting image is practically indistinguishable from the original. These rates are sufficient for the most demanding applications.

6.10 FURTHER READING

Any standard text on information theory covers source-coding theorems and algorithms in detail. Gallager (1968), Blahut (1987), and particularly Cover and Thomas (1991) provide nice and readable treatments of the subject. Our treatment of the Lempel-Ziv algorithm follows that of Cover and Thomas (1991). Berger (1971) is devoted entirely to rate-distortion theory. Jayant and Noll (1984) and Gersho and Gray (1992) examine various quantization and waveform-coding techniques in detail. Gersho and Gray (1992) includes detailed treatment of vector quantization. Analysis-synthesis techniques and linear-predictive coding are treated in books on speech coding such as Markel and Gray (1976), Rabiner and Schafer (1978), and Deller, Proakis, and Hansen (2000). The JPEG standard is described in detail in the book by Gibson, et al. (1998).

Among the original works contributing to the material covered in this chapter, we mention Shannon (1948a, 1959), Huffman (1952), Lloyd (1957), Max (1960), Ziv and Lempel (1978), and Linde, Buzo, and Gray (1980).

PROBLEMS

6.1 A source has an alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ with corresponding probabilities $\{0.1, 0.2, 0.3, 0.05, 0.15, 0.2\}$. Find the entropy of this source. Compare this entropy with the entropy of a uniformly distributed source with the same alphabet.

6.2 Let the random variable X be the output of the source that is uniformly distributed with size N. Find its entropy.

6.3 Show that $H(X) \geq 0$ with equality holding if and only if X is deterministic.

6.4 Let X be a geometrically distributed random variable; i.e.,

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

1. Find the entropy of X.
2. Knowing that $X > K$, where K is a positive integer, what is the entropy of X?

6.5 Let $Y = g(X)$, where g denotes a deterministic function. Show that, in general, $H(Y) \leq H(X)$. When does equality hold?

6.6 An information source can be modeled as a bandlimited process with a bandwidth of 6000 Hz. This process is sampled at a rate higher than the Nyquist rate to provide a guard band of 2000 Hz. It is observed that the resulting samples take values in the set $\mathcal{A} = \{-4, -3, -1, 2, 4, 7\}$ with probabilities 0.2, 0.1, 0.15, 0.05, 0.3, 0.2.

What is the entropy of the discrete-time source in bits/output (sample)? What is the entropy in bits/sec?

6.7 Let X denote a random variable distributed on the set $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ with corresponding probabilities $\{p_1, p_2, \ldots, p_N\}$. Let Y be another random variable defined on the same set but distributed uniformly. Show that

$$H(X) \leq H(Y)$$

with equality if and only if X is also uniformly distributed. (Hint: First prove the inequality $\ln x \leq x - 1$ with equality for $x = 1$, then apply this inequality to $\sum_{n=1}^{N} p_n \ln(\frac{\frac{1}{N}}{p_n})$).

6.8 A random variable X is distributed on the set of all positive integers $1, 2, 3, \ldots$ with corresponding probabilities $p_1, p_2, p_3, \ldots$. We further know that the expected value of this random variable is given to be m; i.e.,

$$\sum_{i=1}^{\infty} i p_i = m$$

Show that among all random variables that satisfy the above condition, the geometric random variable which is defined by

$$p_i = \frac{1}{m}\left(1 - \frac{1}{m}\right)^{i-1} \qquad i = 1, 2, 3, \ldots$$

has the highest entropy. (Hint: Define two distributions on the source, the first one being the geometric distribution given above and the second one an arbitrary distribution denoted by q_i, and then apply the approach of Problem 6.7.)

6.9 Two binary random variables X and Y are distributed according to the joint distribution $p(X = Y = 0) = p(X = 0, Y = 1) = p(X = Y = 1) = \frac{1}{3}$. Compute $H(X)$, $H(Y)$, $H(X \mid Y)$, $H(Y \mid X)$, and $H(X, Y)$.

6.10 Show that if $Y = g(X)$ where g denotes a deterministic function, then $H(Y \mid X) = 0$.

6.11 A memoryless source has the alphabet $\mathcal{A} = \{-5, -3, -1, 0, 1, 3, 5\}$ with corresponding probabilities $\{0.05, 0.1, 0.1, 0.15, 0.05, 0.25, 0.3\}$.

1. Find the entropy of the source.
2. Assume that the source is quantized according to the quantization rule

$$\begin{cases} q(-5) = g(-3) = -4, \\ q(-1) = q(0) = q(1) = 0 \\ q(3) = q(5) = 4 \end{cases}$$

Find the entropy of the quantized source.

6.12 Using both definitions of the entropy rate of a process, prove that for a DMS the entropy rate and the entropy are equal.

with the entropy of the source. (In what base would you compute the logarithms in the expression for the entropy for a meaningful comparison?)

6.26 Design a ternary Huffman code for a source with output alphabet probabilities given by $\{0.05, 0.1, 0.15, 0.17, 0.13, 0.4\}$. (Hint: You can add a dummy source output with zero probability.)

6.27 Find the Lempel-Ziv source code for the binary source sequence

0001001000000110000100000000100000010100001000000110100000000110

Recover the original sequence back from the Lempel-Ziv source code. (Hint: You require two passes of the binary sequence to decide on the size of the dictionary.)

6.28 Using the definition of $H(X)$ and $H(X \mid Y)$ show that

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Now by using the approach of Problem 6.7 show that $I(X; Y) \geq 0$ with equality if and only if X and Y are independent.

6.29 Show that

1. $I(X; Y) \leq \min\{H(X), H(Y)\}$.
2. If $|\mathscr{X}|$ and $|\mathscr{Y}|$ represent the size of sets $\mathscr{X}$ and $\mathscr{Y}$, respectively, then $I(X; Y) \leq \min\{\log|\mathscr{X}|, \log|\mathscr{Y}|\}$.

6.30 Show that $I(X; Y) = H(X)+H(Y)-H(X, Y) = H(Y)-H(Y \mid X) = I(Y; X)$.

6.31 Let X denote a binary random variable with $p(X=0) = 1 - p(X=1) = p$ and let Y be a binary random variable that depends on X through $p(Y=1 \mid X=0) = p(Y=0 \mid X=1) = \epsilon$.

1. Find $H(X)$, $H(Y)$, $H(Y \mid X)$, $H(X, Y)$, $H(X \mid Y)$, and $I(X; Y)$.
2. For a fixed ϵ, which p maximizes $I(X; Y)$?
3. For a fixed p, which ϵ minimizes $I(X; Y)$?

6.32 Show that

$$I(X; YZW) = I(X; Y) + I(X : Z \mid Y) + I(X; W \mid ZY)$$

Can you interpret this relation?

6.33 Let X, Y, and Z be three discrete random variables.

1. Show that if $p(x, y, z) = p(z)p(x \mid z)p(y \mid x)$, we have

$$I(X; Y \mid Z) \leq I(X; Y)$$

2. Show that if $p(x, y, z) = p(x)p(y)p(z \mid x, y)$, then

$$I(X; Y) \leq I(X; Y \mid Z)$$

3. In each case give an example where strict inequality holds.

6.34 Let $\mathscr{X}$ and $\mathscr{Y}$ denote finite sets. We denote the probability vectors on $\mathscr{X}$ by $\mathbf{p}$ and the conditional probability matrices on $\mathscr{Y}$ given $\mathscr{X}$ by $\mathbf{Q}$. Then $I(X; Y)$ can be represented as a function of the probability distribution on $\mathscr{X}$ and the conditional probability distribution on $\mathscr{Y}$ given $\mathscr{X}$ as $I(\mathbf{p}; \mathbf{Q})$.

1. Show that $I(\mathbf{p}; \mathbf{Q})$ is a concave function in $\mathbf{p}$; i.e., for any $\mathbf{Q}$, any $0 \leq \lambda \leq 1$, and any two probability vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ on $\mathscr{X}$, we have

$$\lambda I(\mathbf{p}_1; \mathbf{Q}) + \bar{\lambda} I(\mathbf{p}_2; \mathbf{Q}) \leq I(\lambda \mathbf{p}_1 + \bar{\lambda}\mathbf{p}_2; \mathbf{Q})$$

where $\bar{\lambda} \stackrel{\text{def}}{=} 1 - \lambda$.

2. Show that $I(\mathbf{p}; \mathbf{Q})$ is a convex function in $\mathbf{Q}$; i.e., for any $\mathbf{p}$, any $0 \leq \lambda \leq 1$, and any two conditional probabilities $\mathbf{Q}_1$ and $\mathbf{Q}_2$, we have

$$I(\mathbf{p}; \lambda \mathbf{Q}_1 + \bar{\lambda}\mathbf{Q}_2) \leq \lambda I(\mathbf{p}; \mathbf{Q}_1) + \bar{\lambda} I(\mathbf{p}; \mathbf{Q}_2)$$

(Note: You have to first show that $\lambda \mathbf{p}_1 + \bar{\lambda}\mathbf{p}_2$ and $\lambda \mathbf{Q}_1 + \bar{\lambda}\mathbf{Q}_2$ are a legitimate probability vector and conditional probability matrix, respectively.)

6.35 Let the random variable X be continuous with PDF $f_X(x)$ and let $Y = aX$ where a is a nonzero constant.

1. Show that $h(Y) = \log|a| + h(X)$.
2. Does a similar relation hold if X is a discrete random variable?

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

1. X is an exponential random variable with parameter $\lambda > 0$; i.e.,

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

2. X is a Laplacian random variable with parameter $\lambda > 0$; i.e.,

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

3. X is a triangular random variable with parameter $\lambda > 0$; i.e.,

$$f_X(x) = \begin{cases} \frac{x+\lambda}{\lambda^2}, & -\lambda \leq x \leq 0 \\ \frac{-x+\lambda}{\lambda^2}, & 0 < x \leq \lambda \\ 0, & \text{otherwise} \end{cases}$$

6.37 Generalize the technique developed in Problem 6.7 to continuous random variables and show that for continuous X and Y

1. $h(X \mid Y) \leq h(X)$ with equality if and only if X and Y are independent.
2. $I(X; Y) \geq 0$ with equality if and only if X and Y are independent.

the relation in Equation (9.5.37)

$$e_c + e_d = d_{\min} - 1$$

it can detect up to 3 errors with certainty and 4 or more errors with high probability. If a single error is encountered, it is corrected, if multiple errors are detected, then all 28 symbols are flagged as "unreliable." After deinterleaving, these symbols are passed to the decoder for the code C_1. Decoder C_1 tries single error, or 2 erasure corrections. If it fails, all output symbols are flagged; if there are more 3 or more flags at its input, it copies them to its output.

At the output of the second decoder, the symbol corresponding to "unreliable" positions are filled in by interpolation of the other positions. Using this rather complex encoding–decoding technique together with the signal processing methods, burst errors of up to 12,000 data bits, which correspond to a track length of 7.5 mm on the disc, can be concealed.

9.11 FURTHER READING

The noisy channel-coding theorem, which plays a central role in information theory, and the concept of channel capacity were first proposed and proved by Shannon (1948). For detailed discussion of this theorem and its several variations the reader may refer to standard books on information theory such as Gallager (1968), Blahut (1987), and Cover and Thomas (1991).

Golay (1949), Hamming (1950), Hocquenghem (1959), Bose and Ray-Chaudhuri (1960a,b), and Reed and Solomon (1960) are landmark papers in the development of block codes. Convolutional codes were introduced by Elias (1955) and various methods for their decoding were developed by Wozencraft and Reiffen (1961), Fano (1963), Zigangirov (1966), Viterbi (1967), and Jelinek (1969). Trellis-coded modulation was introduced by Ungerboeck (1982) and later developed by Forney (1988a,b). Product codes were introduced by Elias (1954) and concatenated codes were developed and analyzed by Forney (1966). Berrou, Glavieux, and Thitimajshima (1993) introduced turbo codes. The interested reader is referred to books on coding theory including Berlekamp (1968), Peterson and Weldon (1972), MacWilliams and Sloane (1977), Lin and Costello (1983), and Blahut (1983). The reader is referred to the book by Heegard and Wicker (1999) on turbo codes.

PROBLEMS

9.1 Find the capacity of the channel shown in Figure P-9.1.

9.2 The channel shown in Figure P-9.2 is known as the *binary erasure channel*. Find the capacity of this channel and plot it as a function of ϵ.

9.3 Find the capacity of the cascade connection of n binary-symmetric channels with same crossover probability ϵ. What is the capacity when the number of channels goes to infinity.

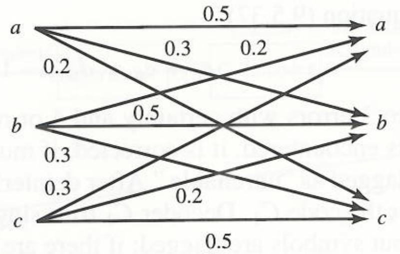

Figure P-9.1

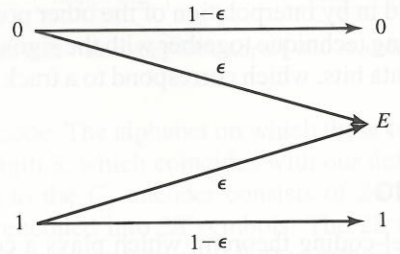

Figure P-9.2

9.4 Using Stirling's approximation $n! \approx n^n e^{-n}\sqrt{2\pi n}$, show that

$$\binom{n}{n\epsilon} \approx 2^{nH_b(\epsilon)}$$

9.5 Show that the capacity of a binary-input, continuous-output AWGN channel with inputs $\pm A$ and noise variance σ^2 (see Example 9.1.2) is given by

$$C = \frac{1}{2} f\left(\frac{A}{\sigma}\right) + \frac{1}{2} f\left(-\frac{A}{\sigma}\right)$$

where

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(u-x)^2/2} \log_2 \frac{2}{1+e^{-2xu}}\, du$$

9.6 The matrix whose elements are the transition probabilities of a channel; i.e., $p(y_i \mid x_j)$'s, is called the channel probability transition matrix. A channel is called *symmetric* if all rows of the channel probability transition matrix are permutations of each other, and all its columns are also permutations of each other. Show that in a symmetric channel the input probability distribution that achieves capacity is a uniform distribution. What is the capacity of this channel?

9.7 Channels 1, 2, and 3 are shown in Figure P-9.7.

1. Find the capacity of channel 1. What input distribution achieves capacity?
2. Find the capacity of channel 2. What input distribution achieves capacity?

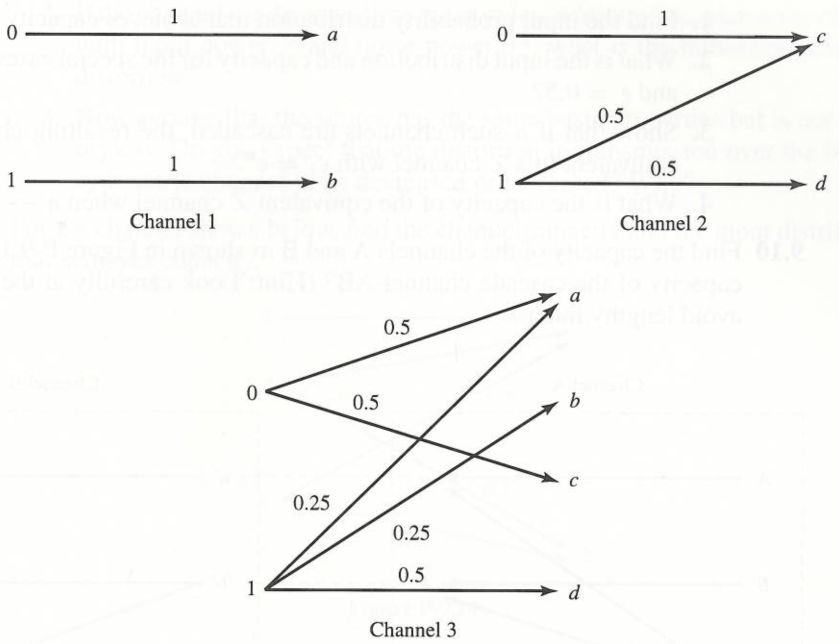

Figure P-9.7

3. Let C denote the capacity of the third channel and C_1 and C_2 represent the capacities of the first and second channel. Which of the following relations holds true and why?

(a) $C < \frac{1}{2}(C_1 + C_2)$.

(b) $C = \frac{1}{2}(C_1 + C_2)$.

(c) $C > \frac{1}{2}(C_1 + C_2)$.

9.8 Let C denote the capacity of a discrete-memoryless channel with input alphabet $\mathscr{X} = \{x_1, x_2, \ldots, x_N\}$ and output alphabet $\mathscr{Y} = \{y_1, y_2, \ldots, y_M\}$. Show that $C \leq \min\{\log M, \log N\}$.

9.9 The channel C is (known as the Z channel) shown in Figure P-9.9.

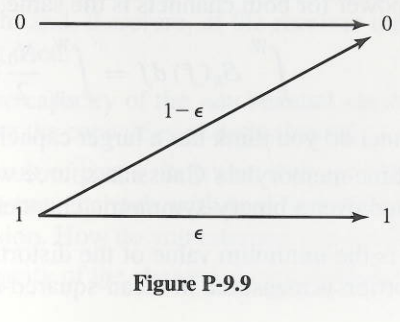

Figure P-9.9

1. Find the input probability distribution that achieves capacity.
2. What is the input distribution and capacity for the special cases $\epsilon = 0$, $\epsilon = 1$ and $\epsilon = 0.5$?
3. Show that if n such channels are cascaded, the resulting channel will be equivalent to a Z channel with $\epsilon_1 = \epsilon^n$.
4. What is the capacity of the equivalent Z channel when $n \to \infty$.

9.10 Find the capacity of the channels A and B as shown in Figure P-9.10. What is the capacity of the cascade channel AB? (Hint: Look carefully at the channels and avoid lengthy math.)

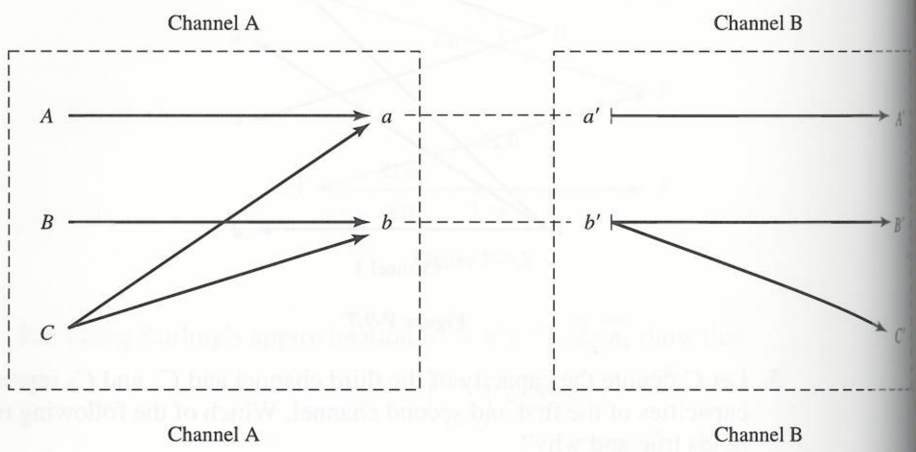

Figure P-9.10

9.11 Find the capacity of an additive white Gaussian noise channel with a bandwidth of 1 MHz, power of 10 W, and noise power-spectral density of $\frac{N_0}{2} = 10^{-9}$ W/Hz.

9.12 Channel C_1 is an additive white Gaussian noise channel with a bandwidth of W, transmitter power of P and noise power-spectral density of $\frac{N_0}{2}$. Channel C_2 is an additive Gaussian noise channel with the same bandwidth and power as channel C_1 but with noise power-spectral density $\mathcal{S}_n(f)$. It is further assumed that the total noise power for both channels is the same, that is

$$\int_{-W}^{W} \mathcal{S}_n(f)\, df = \int_{-W}^{W} \frac{N_0}{2}\, df = N_0 W$$

Which channel do you think has a larger capacity? Give an intuitive reasoning.

9.13 A discrete-time memoryless Gaussian source with mean 0 and variance σ^2 is to be transmitted over a binary-symmetric channel with crossover probability ϵ.

1. What is the minimum value of the distortion attainable at the destination? (Distortion is measured in mean-squared-error.)