$$
\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{H}^{-1^{\mathrm{T}}}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{H}^{-1}\boldsymbol{y}\right)|\boldsymbol{H}|^{-1} \\
&= \frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{xx}}|^{1/2}|\boldsymbol{H}|} \exp\left(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1}\boldsymbol{y}\right)
\end{aligned}
\tag{2.148}
$$

where $\boldsymbol{\Sigma}_{\boldsymbol{yy}}=\boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{H}^{\mathrm{T}}$. Note that a linear transformation of a Gaussian process yields another Gaussian process.

Applications

In this section we consider applications of probability models in signal coding, citation indexing, pattern recognition and noise reduction.

Entropy Coding

Consider a communication system with an alphabet X of M symbols x_1, x_2, .., x_M with probabilities $P_1=P(x_1)$, $P_2=P(x_2)$,,..., $P_M=P(x_M)$. The entropy of X is

$$
H(X) = -\sum_{i=1}^{M} P(x_i)\log P(x_i) \qquad ()
$$

Entropy gives the minimum number of bits required to encode the source. This theoretical minimum is usually approached by encoding N samples of the process simultaneously with K bits where $K/N \geq H(X)$. As N becomes large then for an efficient coder K/N approaches the entropy $H(X)$ of X.
The simplest method to encode an M-valued variable is to use a fixed-length coding scheme that uniformly assigns N binary digits to each of the M values with N=Nint($\log_2 M$), where Nint is nearest integer. *The efficiency of a coding scheme in terms of its entropy is defined as* $H(X)/N$. When $N=H(X)$entropy coding efficiency of the code is $H(X)/N$=1 or 100%. When the source symbols are *not equally probable*, a more efficient method is *entropy encoding*. Entropy coding is a variable length coding procedure which assigns codewords of variable lengths to symbols x_k such that more probable symbols which occur more frequently are assigned shorter codewords and less probable symbols which happen less frequently are

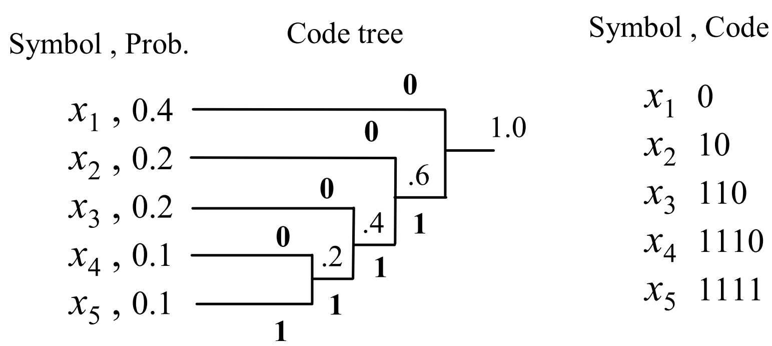

Figure 2.18 Illustration of a tree code and Huffman coding

assigned longer code words. An example of such a code is the Morse Code which dates back to the nineteen century. If the entropy coding is ideal, the bit rate at the output of a uniform M level quantiser can be reduced by an amount of $\log_2 M - H(X)$.

A simple form of entropy coding is the *Huffman Coding* that creates an efficient set of prefix codes for a given text. The ease with which Huffman codes can be created and used makes this code an extremely popular tool for data compression. The procedure consists in arranging source words in decreasing order of probability in a columns. The two lowest probability symbols are combined by drawing a straight line out from each and connecting them. This combination is combined with the next symbol and the procedure is repeated to cover all symbols. Binary codewords are assigned by moving from the root of the tree at the right hand side to left in the tree and assigning a 1 to the lower branch and a 0 to the upper branch where each pair of symbols have been combined.

Example 2.23 Given 5 symbols x_1, x_2, ..., x_5 with probabilities of $P(x_1)$=0.4, $P(x_2)$= $P(x_3)$=0.2 and $P(x_4)$= $P(x_5)$=0.1, design a binary variable length code for this source.
Figure 2.18 illustrates the design of a Huffman code for this source.

The average codeword length = 4 ×0.1 + 4 ×0.1 + 3 × 0.2 + 2 × 0.2 + 1 × 0.4

= 2.2 bits/symbol

The entropy of X is $H(X)$=1.86 bits/symbol. The average code word length of 2.2 bits/symbol is close to the minimum possible value of 1.86 bits/symbol. We can get closer to the minimum average codeword length by encoding pairs of letters or blocks of more than two symbols at a time (with added complexity). The Huffman code has an important *prefix condition* property whereby no codeword is a prefix or an initial part of another codeword. Thus codewords can be readily concatenated (in a comma free fashion) and be uniquely (unambiguously) decoded.

Web Page Citation Ranking and Indexing

Search engines on the world wide web have to sort billions of web pages and websites. A good set of search keywords focus the search on documents and websites which contain the search words. However the problem remains that often the contents of many websites are not of the required quality and there are also misleading websites containing words aimed to attract visitors to increase the hit rates or advertising revenue. For efficient information management the websites and their information content need to be ranked using an objective quality measure. An objective measure of quality of published information on any medium is citation, which for long has been used in academic research. A map of hyperlinks and pointers on the web allows rapid calculation of a web page's rank in terms of citation. Page rank is a good way to prioritise the results of web keyword searches.

Citation Ranking in Web Page Rank Calculation

Search engines usually find many web pages that contain the search keywords. The problem is how to present the web links containing the search keywords in a rank ordered form, such that the rank of a page represents a measure of the quality of information on the page. The relevance of a web page containing the search text string can be determined from the following analysis of the web page:

(a) Page title containing the search words is an indicator of the relevance of the topic of the page, but not of its quality.
(b) Number of times the search words are mentioned in the web page is also an indicator.

(c) Number of citation of the web page from other web pages is an objective indicator of quality as perceived by web users.
(d) Each citation to a web page can be weighted by its importance which itself is a weighted citation of the citation.

The simplest way to rank a web page is to count the total number of citation links pointing to that page and then divide this by the total number of citations links on the web. This method would rank a web page using a simple probability measure defined as the frequency of citation links. However, as with the tradition of academic research, a citation itself need to be weighted by the quality of the source of citation i.e. by the citation ranking of the source itself. A weighted citation gives some approximation of a page's importance or quality, where each source of citation is weighted by its own citation ranking.

Let $PR(A)$ define the page rank for a web page A. Assume that page A has pages $T_1 ... T_n$ pointing to it. Page rank of A can be defined as

$$PR(A) = (1\text{-}d) + d\ (PR(T_1)/C(T_1) + ... + PR(Tn)/C(T_n))$$

where $C(T)$ is defined as the number of links going out of page T. The parameter d is a damping factor which can be set between 0 and 1, usually d is set to 0.85. Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one.

PageRank or $PR(A)$ can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web. PageRank can be thought of as a model of user behaviour. It is assumed there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. And, the d damping factor is the probability at each page the "random surfer" will get bored and request another random page. One important variation is to only add the damping factor d to a single page, or a group of pages. This allows for personalization and can make it nearly impossible to deliberately mislead the system in order to get a higher ranking.

Another intuitive justification is that a page can have a high PageRank if there are many pages that point to it, or if there are some pages that point to it and have a high PageRank. Intuitively, pages that are well cited from many places around the web are worth looking at. Also, pages that have perhaps only one citation from something like the homepage are also generally worth looking at. If a page was not high quality, or was a broken link, it is quite likely that Yahoo's homepage would not link to it. PageRank handles both these cases and everything in between by recursively propagating weights through the link structure of the web.

2.7 Summary

The theory of statistical processes is central to the development of signal processing algorithms. We began this chapter with basic definitions of deterministic signals, random signals and random processes. A random process generates random signals, and the collection of all signals that can be generated by a random process is the space of the process. Probabilistic models and statistical measures, originally developed for random variables, were extended to model random signals. Although random signals are completely described in terms of probabilistic models, for many applications it may be sufficient to characterise a process in terms of a set of relatively simple statistics such as the mean, the autocorrelation function, the covariance and the power spectrum. Much of the theory and application of signal processing is concerned with the identification, extraction, and utilisation of structures and patterns in a signal process. The correlation and its Fourier transform the power spectrum are particularly important because they can be used to identify the patterns in a stochastic process.

We considered the concepts of stationary, ergodic stationary and non-stationary processes. The concept of a stationary process is central to the theory of linear time-invariant systems, and furthermore even non-stationary processes can be modelled with a chain of stationary sub-processes as described in Chapter 5 on hidden Markov models. For signal processing applications, a number of useful pdfs, including the Gaussian, the mixture Gaussian, the Markov and the Poisson process, were considered. These pdf models are extensively employed in the remainder of this book. Signal processing normally involves the filtering or transformation of an input signal to an output signal. We derived general expressions for the pdf of the output of a system in terms of the pdf of the input. We also considered some applications of stochastic processes for modelling random noise such as white noise, clutters, shot noise and impulsive noise.

Bibliography

ANDERSON O.D. (1976) Time Series Analysis and Forecasting. The Box–Jenkins Approach. Butterworth, London.