FUNDAMENTAL OF RANDOM SIGNALS

CONTENTS

REFERENCES: N. S. JAYANT AND PETER NOLL, DIGITAL CODING OF WAVEFORMS: PRINCIPLE AND APPLICATION TO SPEECH AND VIDEO. PHI. 1984. CHAPTER TWO.

1. Random Variables

1.1 Probability distribution and density functions

Let X be a random variable (r.v.), the probability distribution function is:

$$P_X(x) = P\{X \le x\}, \qquad (1)$$

which is the probability of the random variable X less than or equal to x.

If $P_X(x)$ is continuous, then the probability density function (pdf) is:

$$p_X(x) = \frac{dP_X(x)}{dx}. \qquad (2)$$

$P_X(x)$ and $p_X(x)$ satisfy the properties:

$$p_X(x) \ge 0, \forall x; \qquad \int_{-\infty}^{+\infty} p_X(x)dx = 1. \qquad (3a)$$

$$p_X(x) \geq 0, \forall x; \qquad \int_{-\infty}^{+\infty} p_X(x)dx = 1. \tag{3b}$$

$$P_X(-\infty) = 0; P_X(\infty) = 1.. \tag{3c}$$

1.2 Expectation of Random Variables

If $g(x)$ is a function of the r.v., the expectation of the new r.v. $g(X)$ is:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) p_X(x)dx. \tag{4}$$

Mean value:

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x p_X(x)dx. \tag{5}$$

Mean square value:

$$\chi_X^2 = E[X^2] = \int_{-\infty}^{+\infty} x^2 p_X(x)dx. \tag{6}$$

Variance:

$$\sigma_X^2 = E[(X-\mu_X)^2] \\ = \int_{-\infty}^{+\infty} (x-\mu_X)^2 p_X(x)dx \quad (7)$$

n-th moment: $E[X^n]$. **(8)**

n-th central moment: $E[(X-\mu_N)^n]$. **(9)**

The variance measures the spread of the distribution about its mean value.

σ_X is called the standard deviation or rms value:

1.3 Bivariate Distribution

Let X and Y be two r.v.'s whose mean values and variances are denoted by μ_X, σ_X^2, μ_Y, σ_Y^2, respectively.

The joint probability distribution and joint pdf are defined by:

$$P_{XY}(x,y) = P\{X \le x, Y \le y\}. \quad (10)$$

$$p_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} P_{XY}\{x,y\}$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{XY}(x,y)dxdy = 1$$

. **(11)**

The probability that the r.v.'s X and Y are jointly in ranges (a,b) and (c,d) is:

$$P\{a < X \le b, c < Y \le d\} = \int_a^b \int_c^d p_{XY}(x,y)dxdy.$$ **(12)**

The marginal pdf's are given by:

$$p_X = \int_{-\infty}^{+\infty} p_{XY}(x,y)dy \quad p_Y = \int_{-\infty}^{+\infty} p_{XY}(x,y)dy.$$ **(13)**

Statistic Independent and uncorrelated

- The r.v.'s X and Y are said to be **statistically independent** if

$$p_{XY}(x,y) = p_X(x)p_Y(y). \qquad (14)$$

- The expectation of a function of X and Y is:

$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)p_{XY}(x,y)dxdy. \qquad (15)$$

统计上相互独立
可以拆开来计算

The joint moment is given by $E[X^m Y^n]$. It is easy to see that $E[X^m Y^n] = E[X^m]E[Y^n]$ if X and Y are statistically independent.

- X and Y are said to be **uncorrelated** if $E[XY] = E[X]E[Y]$ and **orthogonal** if $E[XY] = 0$.

正交的

Useful properties

- **Mean of the sum of two r.v.'s is the sum of their means:**

$$\mu_{X+Y} = E[X+Y] = E[X] + E[Y] = \mu_X + \mu_Y. \tag{16}$$

- **The power of their sum is:**

$$\chi^2_{X+Y} = E[(X+Y)^2] = E[X^2] + E[Y^2] + 2E[XY]. \tag{17}$$

For uncorrelated r.v.'s, we have:

$$\boxed{\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y.} \tag{18}$$

- **The joint moment:**

$$\sigma_{X+Y} = E[(X-\mu_X)(Y-\mu_Y)]. \tag{19}$$

is called the covariance of the r.v.'s X and Y. The normalized covariance of X and Y.

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \tag{20}$$

is called the correlation coefficient.

1.4 Conditional Distributions

The conditional probability density function of r.v.s X and Y, knowing a single point event $\{Y = y\}$ is

$$p_X(x \mid Y = y) = \frac{p_{XY}(x, y)}{p_Y(y)}, \tag{21}$$

provided that $p_Y(y) \neq 0$. If the r.v.'s are independent then $p_X(x \mid Y = y) = p_X(x)$.

The conditional expectation is the expectation of $g(X)$ knowing that event $Y = y$ has occurred:

$$E[g(X) \mid Y = y] = \int_{-\infty}^{+\infty} g(x) p_X(x \mid Y = y) dx. \tag{22}$$

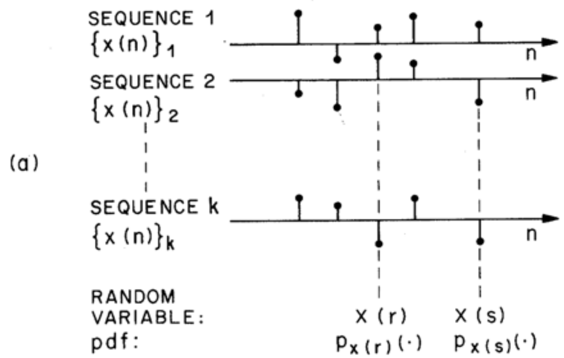
The conditional pdf of a r.v. X given that event $\{x \in I\}$ has occurred is:

$$p_X(x \mid X \in I) = \begin{cases} \dfrac{p_X(x)}{P\{X \in I\}} & X \in I \\ 0 & otherwise \end{cases}. \tag{23}$$

2. Random Processes

- **Communications waveforms are not deterministic - the signal generating rule is either not known or is so complex as to make precise signal description impossible. They are usually modeled as random or stochastic signals.**

Figure 2.11 Random processes in (a) one and (b) two dimensions [Rosenfield and Kak, 1976. Reprinted with permission].



Each sequence from a random process is considered to be a member of an ensemble of signals. Each realization of the random process is chosen randomly from the ensemble by nature.

Stationary Processes

- **At each time instant n, x(n) of the sequence {x(n)} is a particular value of a random variable X(n) which can be described by a set of probability distributions function**

$$P_{x(n)}(x) = P\{X(n) \le x\}. \quad \textbf{(24)}$$

and probability density function (pdf):

$$p_{x(n)}(x) = \frac{dP_{x(n)}(x)}{dx}. \quad \textbf{(25)}$$

- **In case of non-stationary processes, all expectations will be functions of time index n. In wide-sense stationary process, the first two moments of the distribution are independent of time:**

$$\mu_{x(n)} = \mu_{x(n+k)} = \mu_x; \quad \sigma^2_{x(n)} = \sigma^2_{x(n+k)} = \sigma^2_x \quad \textbf{(26)}$$

Hence the first and second-order statistics are independent of time shift. In strict-sense stationary process, none of the statistics is affected by a time shift.

Correlation functions

- **Correlation functions are used to describe the behavior of r.v.'s X(n) and X(m) of a random process.**

Autocovariance function:

$$C_{xx}(m,n) = E[X(m) - \mu_{x(m)})(X(n) - \mu_{x(n)})] \quad (27)$$

Autocorrelation function (acf):

$$R_{xx}(m,n) = E[X(m)X(n)] = C_{xx}(m,n) + \mu_{x(m)}\mu_{x(n)} \quad (28)$$

For zero mean process, the two functions are identical. Both the acf and autocovariance functions depend on the time variables m and n (or m and time lag k = n-m).

For wide-sense stationary signals, the joint pdf $p_{x(m)x(n)}(x, y)$ depends only on time shift implying that the acf depends only on $k = n - m$. We may then write:

$$C_{xx}(k) = E[(X(n) - \mu_x)(X(n+k) - \mu_x)]. \quad (29)$$

$$R_{xx}(k) = E[(X(n)X(n+k)]. \quad (30)$$

Some properties of the acf are;

$$R_{xx}(k) = R_{xx}(-k); \qquad |R_{xx}(k)| \leq R_{xx}(0). \quad (31)$$

$$R_{xx}(0) = \chi_x^2 = \sigma_x^2 + \mu_x^2. \quad (32)$$

Variance normalized acf:

$$\rho_k = \rho_{xx}(k) = \frac{R_{xx}(k)}{R_{xx}(0)}; \rho_{xx}(0) = 1. \quad (33)$$

2.2 Power spectral density of random processes

Suppose that $\{X(n)\}$ is a discrete-time wide-sense stationary process with acf $R_{xx}(k)$. The power spectral density function (psd) is defined to be the Fourier transform of the acf of the process (Wiener-Khinchine theorem):

$$R_{xx}(k) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} S_{xx}(e^{j\omega})e^{jk\omega}d\omega. \tag{34}$$

$$S_{xx}(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} R_{xx}(k)e^{-jk\omega}. \tag{35}$$

The function $S_{xx}(e^{j\omega})$ exists if $R_{xx}(k)$ is absolutely summable. Since the acf is real and even, the PSD is real, even and positive. Specifically, the power of the process in the frequency range $(\omega_k - \Delta\omega/2, \omega_k + \Delta\omega/2)$ is approximately $S_{xx}(e^{j\omega})\Delta\omega/\pi$.

The cross-spectral density of two jointly stationary processes $\{X(n)\}$ and $\{Y(n)\}$ is

$$S_{xy}(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} R_{xy}(k) e^{-jk\omega}. \tag{36}$$

Note that the csd function is not necessarily even or real, or positive.

2.3 Power spectral density of random processes

Consider a wide-sense stationary random process $\{X(n)\}$ applied to the input of a linear, time-invariant and stable system with impulse response $h(n)$ and transfer function $H(e^{j\omega})$. The mean and acf of the output process $\{Y(n)\}$ are then given by:

$$\mu_y = \mu_x H(0). \tag{37}$$

$$\begin{aligned} R_{yy}(k) &= R_{xx}(k) * h(k) * h(-k) \\ &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h(i)h(j)R_{xx}(|k-i+j|) \end{aligned}. \tag{38}$$

Since the mean of $\{Y(n)\}$ is constant and its acf depends solely on time displacement k, $\{Y(n)\}$ is also wide-sense stationary.

The PSD of $\{Y(n)\}$ is therefore: $S_{yy}(e^{j\omega}) = S_{xx}(e^{j\omega}) \, | H(e^{j\omega}) |^2$. **(39)**

2.4 Optimum linear predictions

Linear prediction is useful to removing the linear redundancy found in random processes. Consider an N order one-step ahead linear predictor:

$$\hat{x}(n) = \sum_{j=1}^{N} h_j x(n-j). \tag{40}$$

The filter coefficients can be obtained by minimizing the following mean square cost function:

$$\sigma_d^2 = E[d^2(n)] = E[(x(n) - \hat{x}(n))^2]. \tag{41}$$

Differentiating wrt h_i and setting the derivatives to zero, we obtain:

$$\begin{aligned} \frac{\partial \sigma_d^2}{\partial h_i} = 0,\ i = 1,2,\cdots,N \ \Leftrightarrow\ & E[2\{x(n) - \hat{x}_{opt}(n)\} \frac{\partial}{\partial h_i}\{-\hat{x}_{opt}(n)\}] \\ & = E[2\{x(n) - \hat{x}_{opt}(n)\} x(n-i)] = 0 \end{aligned}. \tag{42}$$

Thus the minimum error $x(n)-\hat{x}_{opt}(n)$ must be orthogonal to all the data used in the prediction, which is the orthogonal principle. Expanding equation (42) gives:

$$R_{xx}(k)=\sum_{j=1}^{N} h_{j,opt} R_{xx}(j-k); \qquad k=1,2,\cdots,N. \tag{43}$$

or in matrix form

$$\boldsymbol{r}_{xx}=\boldsymbol{R}_{xx}\boldsymbol{h}_{opt} \text{ or } \boldsymbol{h}_{opt}=\boldsymbol{R}_{xx}^{-1}\boldsymbol{r}_{xx} \quad \text{(if } \boldsymbol{R}_{xx} \text{ is nonsingular).} \tag{44}$$

where $\boldsymbol{r}_{xx}^{T}=\{R_{xx}(i)\}$; $\boldsymbol{R}_{xx}=\{R_{xx}(|i-j|)\}$; $\quad i,j=1,2,\cdots,N$.

The equations are called normal equations, Yule-Walker prediction equations or Wiener-Hopf equations. The minimum mse is given by:

$$\min\{\sigma_d^2\}=\sigma_x^2-\boldsymbol{h}_{opt}^{T}\boldsymbol{r}_{xx}. \tag{45}$$

Since $\boldsymbol{R}_{xx}$ is symmetric Toeplitz, it can be solved using the Levinson-Durbin recursion.

If the waveform sources are modeled as the output a linear filter with a white noise or innovations process $\{Z(n)\}$ as its input, the asymptotic value of minimum mse as $N \to \infty$ is merely given by the variance σ_z^2. The optimal prediction error filter is in a fact a whitening filter for the waveform at the input to a coder, and its transfer function is the inverse of the input psd (c.f. section 2.4).

2.4 AR, MA, ARMA processes

- **Interesting classes of discrete-time linear processes can be generated by passing constant-psd or white-noise sequences through a linear time-invariant filter with impulse response $h(n)$.**
- **A wide sense stationary zero mean white noise process $\{Z(n)\}$ is defined by:**

$$S_{zz}(e^{j\omega}) = \sigma_z^2 . \tag{46a}$$

$$R_{zz}(k) = E[Z(n)Z(n+k)] = \sigma_z^2 \delta(k) . \tag{46b}$$

In general, the transfer function of an LTI system can be written as:

$$H(z) = \frac{A(z)}{B(z)} = \frac{\sum_{j=0}^{M} a_j z^{-j}}{1 - \sum_{j=1}^{N} b_j z^{-j}} . \tag{47}$$

The output is given by:

$$X(n) = \sum_{j=0}^{M} a_j Z(n-j) + \sum_{j=1}^{N} b_j X(n-j). \qquad (48)$$

If the filter is stable, $X(n)$ is a wide-sense stationary process with acf:

$$R_{XX}(n) = R_{ZZ}(n) * h(n) * h(-n) \qquad (49)$$

and z-transform:

$$S_{xx}(e^{j\omega}) = \sigma_z^2 H(z) H(z^{-1}). \qquad (50)$$

If the filter is an all-pole filter, the process $X(n)$ is called an autoregressive (AR) process. If the filter is an all-zero filter, the process $X(n)$ is moving average (MA) process. The process described by equation (48) is called an ARMA process.