15 Random Vector

In Section 11.2, we have introduced the way to deal with more than two random variables. In particular, we introduce the concepts of joint pmf:

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right]$$

and joint pdf:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

of a collection of random variables.

Definition 15.1. You may notice that it is tedious to write the n-tuple $(X_1, X_2, \ldots, X_n)$ every time that we want to refer to this collection of random variables. A more convenient notation uses a ***column vector*** $\mathbf{X}$ to represent all of them at once, keeping in mind that the ith component of $\mathbf{X}$ is the random variable X_i. This allows us to express

(a) $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ as $p_{\mathbf{X}}(\mathbf{x})$ and

(b) $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ as $f_{\mathbf{X}}(\mathbf{x})$.

When the random variables are separated into two groups, we may label those in a group as $X_1, X_2, \ldots, X_n$ and those in another group as $Y_1, Y_2, \ldots, Y_m$. In which case, we can express

(a) $p_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1, \ldots, x_n, y_1, \ldots, y_m)$ as $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ and

(b) $f_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1, \ldots, x_n, y_1, \ldots, y_m)$ as $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$.

Definition 15.2. Random vectors $\mathbf{X}$ and $\mathbf{Y}$ are ***independent*** if and only if

(a) Discrete: $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})$.

(b) Continuous: $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y})$.

Definition 15.3. A random vector $\mathbf{X}$ contains many random variables. Each of these random variables has its own expected value. We can represent the expected values of all these random variables in the form of a vector as well by using the notation $\mathbb{E}\left[\mathbf{X}\right]$. This is a vector whose ith component is $\mathbb{E}X_i$.

- In other words, the expectation $\mathbb{E}[\mathbf{X}]$ of a random vector $\mathbf{X}$ is defined to be the vector of expectations of its entries.
- $\mathbb{E}[\mathbf{X}]$ is usually denoted by $\mu_{\mathbf{X}}$ or $m_{\mathbf{X}}$.

Definition 15.4. Recall that a random vector is simply a vector containing random variables as its components. We can also talk about ***random matrix*** which is simply a matrix whose entries are random variables. In which case, we define the expectation of a random matrix to be a matrix whose entries are expectation of the corresponding random variables in the random matrix.

Example 15.5.

$$\mathbb{E}\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \begin{pmatrix} \mathbb{E}X \\ \mathbb{E}Y \end{pmatrix}$$

and

$$\mathbb{E}\left[\begin{pmatrix} X & W \\ Y & Z \end{pmatrix}\right] = \begin{pmatrix} \mathbb{E}X & \mathbb{E}W \\ \mathbb{E}Y & \mathbb{E}Z \end{pmatrix}$$

15.6. For non-random matrix A, B, C and a random vector $\mathbf{X}$,

$$\mathbb{E}[A\mathbf{X}B + C] = A(\mathbb{E}\mathbf{X})B + C.$$

Correlation and covariance are important quantities that capture linear dependency between two random variables. When we have many random variables, there are many possible pairs to find correlation $\mathbb{E}[X_i X_j]$ and covariance $\operatorname{Cov}[X_i, X_j]$. All of the correlation values can be expressed at once using the correlation matrix.

Definition 15.7. The ***correlation matrix*** $R_{\mathbf{X}}$ of a random (column) vector $\mathbf{X}$ is defined by

$$R_{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right].$$

Note that it is symmetric and that the ij-entry of $R_{\mathbf{X}}$ is simply $\mathbb{E}[X_i X_j]$.

Example 15.8. Consider $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

$$\begin{aligned} R_{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] &= \mathbb{E}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\begin{pmatrix} X_1 & X_2 \end{pmatrix}\right] \\ &= \mathbb{E}\left[\begin{pmatrix} X_1^2 & X_1X_2 \\ X_1X_2 & X_2^2 \end{pmatrix}\right] = \begin{pmatrix} \mathbb{E}\left[X_1^2\right] & \mathbb{E}[X_1X_2] \\ \mathbb{E}[X_1X_2] & \mathbb{E}\left[X_2^2\right] \end{pmatrix} \end{aligned}$$

Definition 15.9. Similarly, all of the covariance values can be expressed at once using the covariance matrix. The covariance matrix $C_{\mathbf{X}}$ of a random vector $\mathbf{X}$ is defined as

$$\begin{aligned} C_{\mathbf{X}} &= \mathbb{E}\left[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^T\right] = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] - (\mathbb{E}\mathbf{X})(\mathbb{E}\mathbf{X})^T \\ &= R_{\mathbf{X}} - (\mathbb{E}\mathbf{X})(\mathbb{E}\mathbf{X})^T. \end{aligned}$$

Note that it is symmetric and that the ij-entry of $C_{\mathbf{X}}$ is simply $\operatorname{Cov}[X_i, X_j]$.

- In some references, $\Lambda_{\mathbf{X}}$ or $\Sigma_{\mathbf{X}}$ is used instead of $C_{\mathbf{X}}$.

15.10. Properties of covariance matrix:

(a) For i.i.d. X_i each with variance σ^2, $C_{\mathbf{X}} = \sigma^2 I$.

(b) $\operatorname{Cov}[A\mathbf{X}+b] = AC_{\mathbf{X}}A^T$.

In addition to the correlations and covariances of the elements of one random vector, it is useful to refer to the correlations and covariances of elements of two random vectors.

Definition 15.11. If $\mathbf{X}$ and $\mathbf{Y}$ are both random vectors (not necessarily of the same dimension), then their cross-correlation matrix is

$$R_{\mathbf{XY}} = \mathbb{E}\left[\mathbf{X}\mathbf{Y}^T\right].$$

and their cross-covariance matrix is

$$C_{\mathbf{XY}} = \mathbb{E}\left[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{Y}-\mathbb{E}\mathbf{Y})^T\right].$$

Example 15.12. ***Jointly Gaussian random vector*** $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \Lambda)$:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Lambda)}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\Lambda^{-1}(\mathbf{x}-\mathbf{m})}.$$

(a) $\mathbf{m} = \mathbb{E}\mathbf{X}$ and $\Lambda = C_{\mathbf{X}} = \mathbb{E}\left[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^T\right]$.

(b) For bivariate normal, $X_1 = X$ and $X_2 = Y$. We have

$$\Lambda = \begin{pmatrix} \sigma_X^2 & \operatorname{Cov}[X,Y] \\ \operatorname{Cov}[X,Y] & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

16 Introduction to Stochastic Processes (Random Processes)

A random process consider an infinite collection of random variables. These random variables are usually indexed by time. So, the obvious notation for random process would be $X(t)$. As in the signals-and-systems class, time can be discrete or continuous. When time is discrete, it may be more appropriate to use $X_1, X_2, \dots$ or $X[1], X[2], X[3], \dots$ to denote a random process.

Example 16.1. Sequence of results (0 or 1) from a sequence of Bernoulli trials is a discrete-time random process.

16.2. Two perspectives:

(a) We can view a random process as a collection of many random variables indexed by t.

(b) We can also view a random process as the outcome of a random experiment, where the outcome of each trial is a deterministic waveform (or sequence) that is a function of t.

    The collection of these functions is known as an ***ensemble***, and each member is called a ***sample function***.

Example 16.3. Gaussian Random Processes: A random process $X(t)$ is Gaussian if for all positive integers n and for all $t_1, t_2, \dots, t_n$, the random variables $X(t_1), X(t_2), \dots, X(t_n)$ are jointly Gaussian random variables.

16.4. Formal definition of random process requires going back to the probability space $(\Omega, \mathcal{A}, P)$.

Recall that a random variable X is in fact a deterministic function of the outcome ω from Ω. So, we should have been writing it as $X(\omega)$. However, as we get more familiar with the concept of random variable, we usually drop the "(ω)" part and simply refer to it as X.

For random process, we have $X(t, \omega)$. This two-argument expression corresponds to the two perspectives that we have just discussed earlier.

(a) When you fix the time t, you get a random variable from a random process.

(b) When you fix ω, you get a deterministic function of time from a random process.

As we get more familiar with the concept of random processes, we again drop the ω argument.

Definition 16.5. A ***sample function*** $x(t, \omega)$ is the time function associated with the outcome ω of an experiment.

Example 16.6 (Randomly Scaled Sinusoid)**.** Consider the random process defined by

$$X(t) = A \times \cos(1000t)$$

where A is a random variable. For example, A could be a Bernoulli random variable with parameter p.

This is a good model for a one-shot digital transmission via amplitude modulation.

(a) Consider the time $t = 2$ ms. $X(t)$ is a random variable taking the value $1\cos(2) = -0.4161$ with probability p and value $0\cos(2) = 0$ with probability $1 - p$.

If you consider $t = 4$ ms. $X(t)$ is a random variable taking the value $1\cos(4) = -0.6536$ with probability p and value $0\cos(4) = 0$ with probability $1 - p$.

(b) From another perspective, we can look at the process $X(t)$ as two possible waveforms $\cos(1000t)$ and 0. The first one happens with probability p; the second one happens with probability $1 - p$. In this view, notice that each of the waveforms is not random. They are deterministic. Randomness in this situation is associated not with the waveform but with the uncertainty as to which waveform will occur in a given trial.

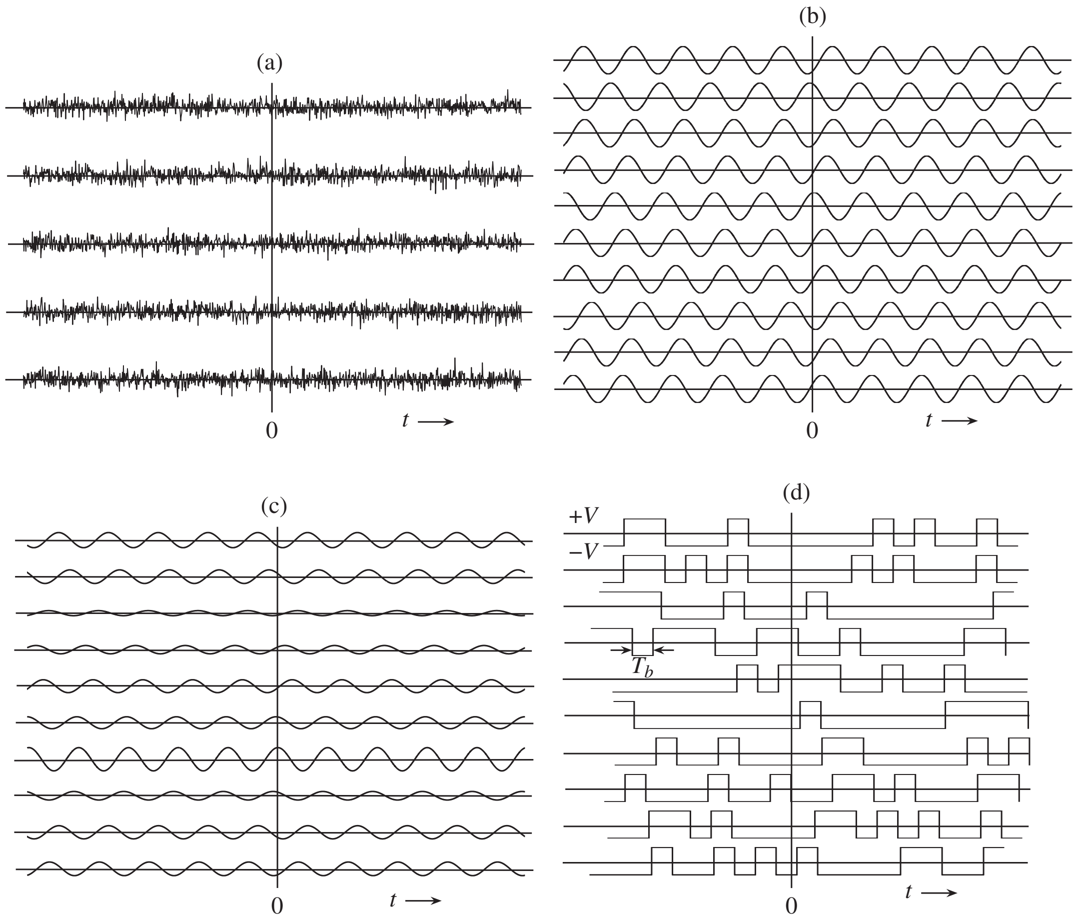

Figure 30: Typical ensemble members for four random processes commonly encountered in communications: (a) thermal noise, (b) uniform phase (encountered in communication systems where it is not feasible to establish timing at the receiver.), (c) Rayleigh fading process, and (d) binary random data process (which may represent transmitted bits 0 and 1 that are mapped to +V and V (volts)). [16, Fig. 3.8]

Definition 16.7. At any particular time t, because we have a random variable, we can also find its expected value. The function $m_X(t)$ captures these expected values as a deterministic function of time:

$$m_X(t) = \mathbb{E}[X(t)].$$

16.1 Autocorrelation Function and WSS

One of the most important characteristics of a random process is its autocorrelation function, which leads to the spectral information of the random process. The frequency content process depends on the rapidity of the amplitude change with time. This can be measured by correlating the values of the process at two time instances t_l and t_2.

Definition 16.8. ***Autocorrelation Function***: The autocorrelation function $R_X(t_1, t_2)$ for a random process $X(t)$ is defined by

$$R_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)].$$

Example 16.9. The random process $x(t)$ is a slowly varying process compared to the process $y(t)$ in Figure 31. For $x(t)$, the values at t_l and t_2 are similar; that is, have stronger correlation. On the other hand, for $y(t)$, values at t_l and t_2 have little resemblance, that is, have weaker correlation.

Example 16.10 (Randomly Phased Sinusoid)**.** Consider a random process

$$X(t) = 5\cos(7t + \Theta)$$

where Θ is a uniform random variable on the interval $(0, 2\pi)$.

$$\begin{aligned} m_X(t) = \mathbb{E}[X(t)] &= \int_{-\infty}^{+\infty} 5\cos(7t + \theta) f_\Theta(\theta) d\theta \\ &= \int_0^{2\pi} 5\cos(7t + \theta) \frac{1}{2\pi} d\theta = 0. \end{aligned}$$

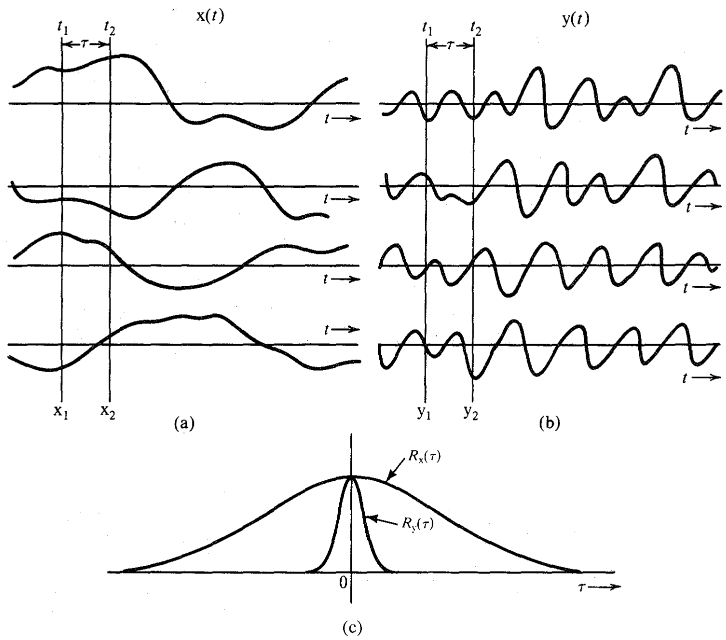

Figure 31: Autocorrelation functions for a slowly varying and a rapidly varying random process [13, Fig. 11.4]

and

$$\begin{aligned} R_X(t_1,t_2) &= \mathbb{E}\left[X(t_1)X(t_2)\right] \\ &= \mathbb{E}\left[5\cos(7t_1+\Theta)\times 5\cos(7t_2+\Theta)\right] \\ &= \frac{25}{2}\cos\left(7(t_2-t_1)\right). \end{aligned}$$

Definition 16.11. A random process whose statistical characteristics do not change with time is classified as a ***stationary*** random process. For a stationary process, we can say that a shift of time origin will be impossible to detect; the process will appear to be the same.

Example 16.12. The random process representing the temperature of a city is an example of a ***nonstationary*** process, because the temperature statistics (mean value, for example) depend on the time of the day.

On the other hand, the noise process is stationary, because its statistics (the mean ad the mean square values, for example) do not change with time.

16.13. In general, it is not easy to determine whether a process is stationary. In practice, we can ascertain stationary if there is no change in the signal-generating mechanism. Such is the case for the noise process.

A process may not be stationary in the strict sense. A more relaxed condition for stationary can also be considered.

Definition 16.14. A random process $X(t)$ is ***wide-sense stationary (WSS)*** if

(a) $m_X(t)$ is a constant

(b) $R_X(t_1, t_2)$ depends only on the time difference $t_2 - t_1$ and does not depend on the specific values of t_1 and t_2.

In which case, we can write the correlation function as $R_X(\tau)$ where $\tau = t_2 - t_1$.

- One important consequence is that $\mathbb{E}\left[X^2(t)\right]$ will be a constant as well.

Example 16.15. The random process defined in Example 16.9 is WSS with

$$R_X(\tau) = \frac{25}{2}\cos(7\tau).$$

16.16. Most information signals and noise sources encountered in communication systems are well modeled as WSS random processes.

Example 16.17. White noise process is a WSS process $N(t)$ whose

(a) $\mathbb{E}[N(t)] = 0$ for all t and

(b) $R_N(\tau) = \frac{N_0}{2}\delta(\tau)$.

See also 16.23 for its definition.

- Since $R_N(\tau) = 0$ for $\tau \neq 0$, any two different samples of white noise, no matter how close in time they are taken, are uncorrelated.

Example 16.18. [Thermal noise] A statistical analysis of the random motion (by thermal agitation) of electrons shows that the autocorrelation of thermal noise $N(t)$ is well modeled as

$$R_N(\tau) = kTG\frac{e^{-\frac{\tau}{t_0}}}{t_0} \text{ watts,}$$

where k is Boltzmann's constant ($k = 1.38 \times 10^{-23}$ joule/degree Kelvin), G is the conductance of the resistor (mhos), T is the (ambient) temperature in degrees Kelvin, and t_0 is the statistical average of time intervals between collisions of free electrons in the resistor, which is on the order of 10^{-12} seconds. [16, p. 105]

16.2 Power Spectral Density (PSD)

An electrical engineer instinctively thinks of signals and linear systems in terms of their frequency-domain descriptions. Linear systems are characterized by their frequency response (the transfer function), and signals are expressed in terms of the relative amplitudes and phases of their frequency components (the Fourier transform). From the knowledge of the input spectrum and transfer function, the response of a linear system to a given signal can be obtained in terms of the frequency content of that signal. This is an important procedure for deterministic signals. We may wonder if similar methods may be found for random processes.

In the study of stochastic processes, the power spectral density function, $S_X(f)$, provides a frequency-domain representation of the time structure of $X(t)$. Intuitively, $S_X(f)$ is the expected value of the squared magnitude of the Fourier transform of a sample function of $X(t)$.

You may recall that not all functions of time have Fourier transforms. For many functions that extend over infinite time, the Fourier transform does not exist. Sample functions $x(t)$ of a stationary stochastic process $X(t)$ are usually of this nature. To work with these functions in the frequency domain, we begin with $X_T(t)$, a truncated version of $X(t)$. It is identical to $X(t)$ for $-T \leq t \leq T$ and 0 elsewhere. We use $\mathcal{F}\{X_T\}(f)$ to represent the Fourier transform of $X_T(t)$ evaluated at the frequency f.

Definition 16.19. Consider a WSS process $X(t)$. The ***power spectral density*** (PSD) is defined as

$$\begin{aligned} S_X(f) &= \lim_{t\to\infty} \frac{1}{2T}\mathbb{E}\left[\left|\mathcal{F}\{X_T\}(f)\right|^2\right] \\ &= \lim_{t\to\infty} \frac{1}{2T}\mathbb{E}\left[\left|\int_{-T}^{T} X(t)e^{-j2\pi ft}dt\right|^2\right] \end{aligned}$$

We refer to $S_X(f)$ as a density function because it can be interpreted as the amount of power in $X(t)$ in the small band of frequencies from f to $f + df$.

16.20. ***Wiener-Khinchine theorem***: the PSD of a WSS random process is the Fourier transform of its autocorrelation function:

$$S_X(f) = \int_{-\infty}^{+\infty} R_X(\tau)e^{-j2\pi f\tau}d\tau$$

and

$$R_X(\tau) = \int_{-\infty}^{+\infty} S_X(f)e^{j2\pi f\tau}df.$$

One important consequence is

$$R_X(0) = \mathbb{E}\left[X^2(t)\right] = \int_{-\infty}^{+\infty} S_X(f)df.$$

Example 16.21. For the thermal noise in Example 16.18, the corresponding PSD is $S_N(f) = \frac{2kTG}{1+(2\pi ft_0)^2}$ watts/hertz.

16.22. Observe that the thermal noise's PSD in Example 16.21 is approximately flat over the frequency range 0–10 gigahertz. As far as a typical communication system is concerned we might as well let the spectrum be flat from 0 to ∞, i.e.,

$$S_N(f) = \frac{N_0}{2} \text{ watts/hertz,}$$

where N_0 is a constant; in this case $N_0 = 4kTG$.

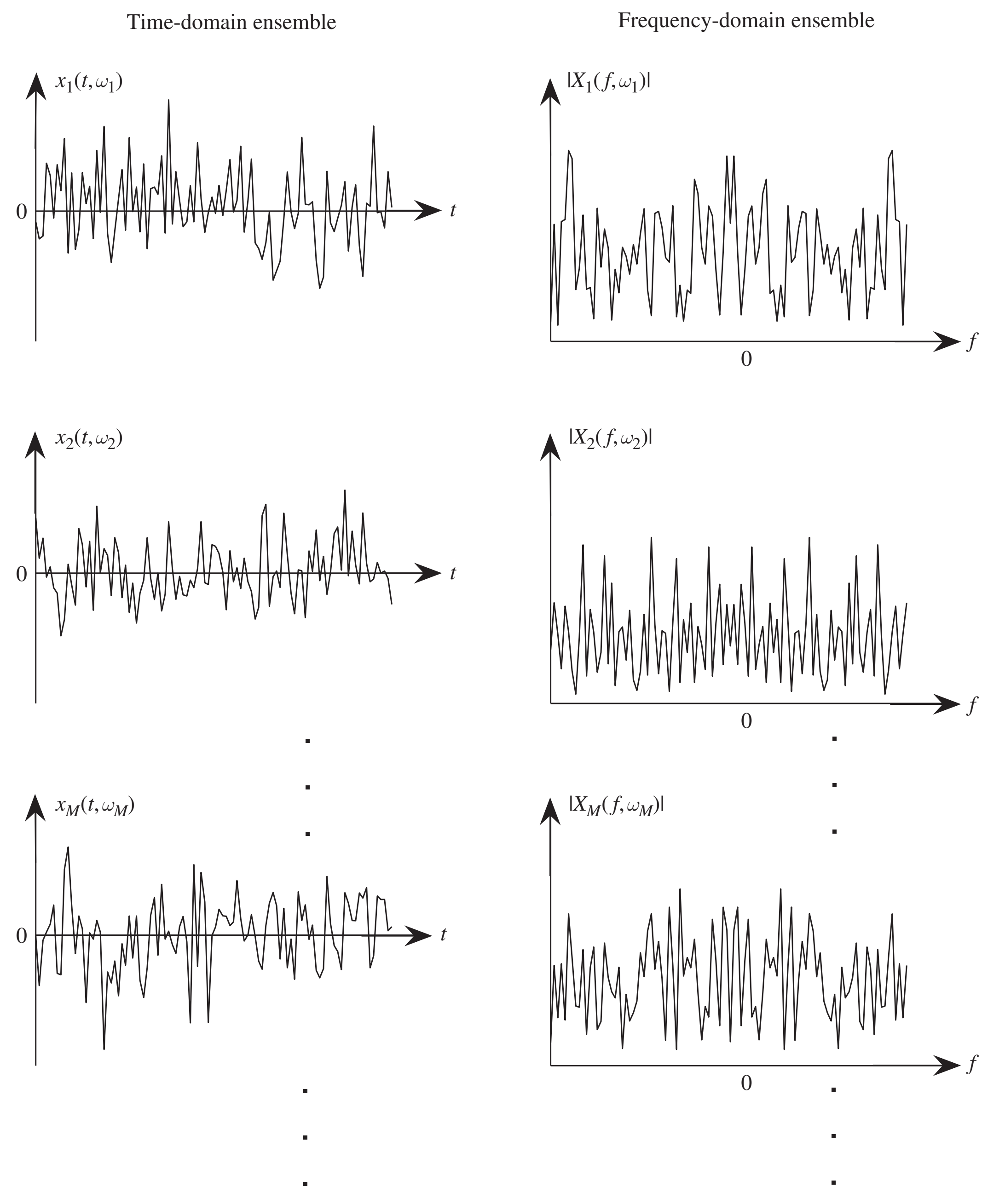

Figure 32: Fourier transforms of member functions of a random process. For simplicity, only the magnitude spectra are shown. [16, Fig. 3.9]

Definition 16.23. Noise that has a uniform spectrum over the entire frequency range is referred to as **white noise**. In particular, for white noise,

$$S_N(f) = \frac{N_0}{2} \text{ watts/hertz,}$$

- The factor 2 in the denominator is included to indicate that $S_N(f)$ is a two-sided spectrum.
- The adjective "white" comes from white light, which contains equal amounts of all frequencies within the visible band of electromagnetic radiation.
- The average power of white noise is obviously infinite.
  (a) White noise is therefore an abstraction since no physical noise process can truly be white.
  (b) Nonetheless, it is a useful abstraction.
    - The noise encountered in many real systems can be assumed to be approximately white.
    - This is because we can only observe such noise after it has passed through a real system, which will have a finite bandwidth. Thus, as long as the bandwidth of the noise is significantly larger than that of the system, the noise can be considered to have an infinite bandwidth.
    - As a rule of thumb, noise is well modeled as white when its PSD is flat over a frequency band that is 35 times that of the communication system under consideration. [16, p 105]

Theorem 16.24. When we input $X(t)$ through an LTI system whose frequency response is $H(f)$. Then, the PSD of the output $Y(t)$ will be given by

$$S_Y(f) = S_X(f)|H(f)|^2.$$

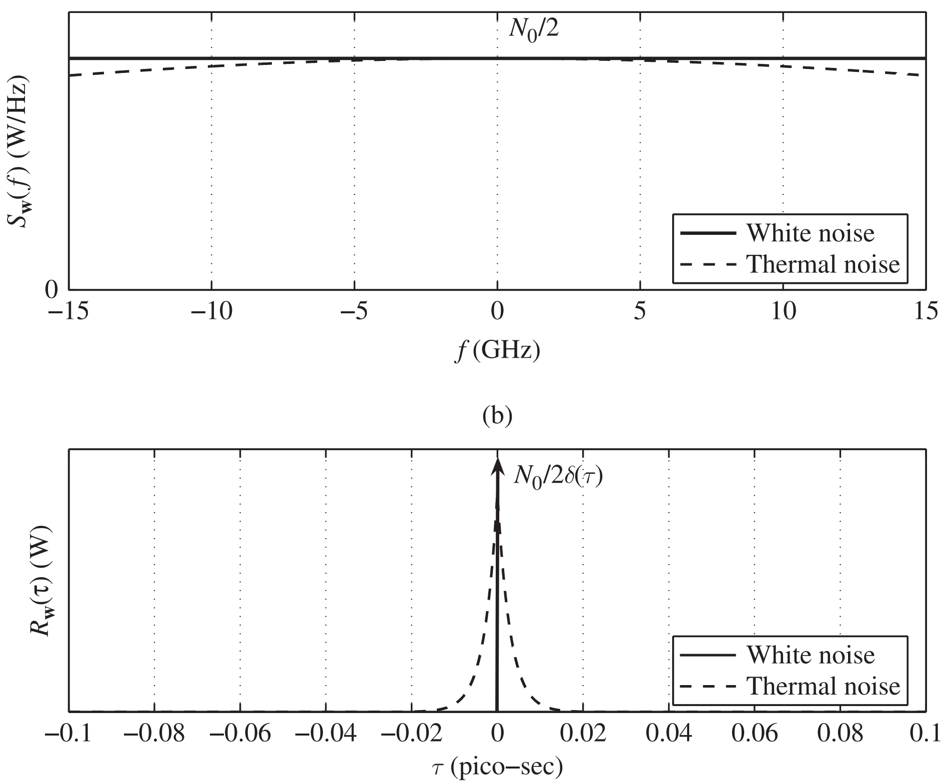

Figure 33: (a) The PSD ($S_N(f)$), and (b) the autocorrelation ($R_N(\tau)$) of noise. (Assume $G = 1/10$ (mhos), $T = 298.15$ K, and $t_0 = 3 \times 10^{-12}$ seconds.) [16, Fig. 3.11]