Random Processes

A random process (RP) or stochastic process is a function that maps all elements of a sample space into a collection or ensemble of time functions called sample functions. The term sample function is used for the time function corresponding to a particular realization of the random process. This is similar to the designation of outcome for a particular realization of a random variable.

The value a random process at any given time cannot be predicted in advance. If a process is not random it is called non-random or deterministic.

As in the treatment of random variables, it is convenient to consider a particular realization of a random process to be determined by a random selection of an element from a sample space $S$. That is, a particular element $s \in S$ is selected according to some random choice and that leads to a particular realization of the RP. To represent the time dependance and random dependance, RP is written as $X(t, s)$ with $t$ representing time and $s$ the randomly chosen element of $S$. As was the case for random variables, for which indicating the dependance on $s$ was often suppressed i.e. $X$ used instead of $X(s)$, a random process will often be written as $X(t)$ instead of $X(t, s)$.

A precise definition of a RP is as follows:

(a) Let $S$ be a non-empty set.
(b) Let $P()$ be a probability measure defined over subsets of $S$, and
(c) To each $s \in S$ let there correspond a time function $X(t, s)$.

Then this probability system is called a RP.

The collection of all possible realizations $X(t, s) : s \in S = \{X(t, s_1), X(t, s_2), ....\}$ is called the ensemble of functions in the RP, where the elements in the set are the sample functions. (As described here, $S$ is countable, but it can be uncountable). Sometimes it is useful to denote the possible values of $t$ as $t \in T$. The ensemble of functions in the random process would then be given by $\{X(t, s); t \in T, s \in S\}$.

Example: A sine wave RP given as

$$X(t) = V \sin(\Omega t + \theta)$$

where amplitude $V$ may be random (AM), frequency $\Omega$ may be random (FM) or phase $\theta$ (PM) may be random or a combination. For the example above, with $\theta \in [0, 2\pi]$ the only RV, there are an infinite number of sample functions. The RP for a simple binary communication $X(t) = \cos(\omega_o t + \theta)$ with $\theta = 0$ or $\pi$, there are only 2 sample functions, $\cos(\omega_o t)$ and $\cos(\omega_o + \pi)$. Another common example is the noise waveform.
Since a RP is a function of two variables, \( t \) and \( s \), either or both of these may be chosen to be fixed. So, the different descriptions of a RP are:

1. \( X(t,s) = X(t) \) is a RP.
2. \( X(t_j,s) = X(s) \) is a RV.
3. \( X(t,s_j) = x(t) \) is a deterministic function of time or sample function.
4. \( X(t_j,s_k) = x \) is a real number.

Both parameters \( t, s \) may be continuous or discrete. We can have a discrete RP, (\( X \) takes on discrete values), a continuous RP or a discrete-time RP \( t \in T \) where \( \{T = t_1, t_2, \ldots t_N\} \) which leads to a random vector.

The complete statistical description of a RP can be infinitely complex. In general, the probability density function of \( X(t_i) \) depends on the value of \( t_i \). If \( X(t) \) is sampled \( N \)-times, then \( X^T = (X(t_1), X(t_2), \ldots, X(t_N) \) is a random vector with joint density function (or joint probability mass function) that depends on \( t_1, t_2, \ldots, t_N \). A suitable description can in theory be given by describing the joint probability density function of \( X \) or joint probability mass function for all \( N \) and for all possible choices of \( t_1, t_2, \ldots, t_N \). Such general characterizations are difficult to analyze. So various simplifying assumptions are typically made. RP specified in several ways:

1. Processes for which the rule for determining the density function is directly stated: e.g. Gaussian RP.
2. Processes consisting of deterministic time functions with parameters that are RVs, e.g. the sine wave RP described earlier.
3. Operations on known RP e.g. filtering a noise RP.
4. Specification of the probability of a finite number of sample functions.

The most common simplifying assumption on RP is that they satisfy some type of stationarity. This is somewhat like steady state. It implies that the the statistics are, in some sense, independent of the absolute values of time. This does not imply that the joint statistics of, say, \( X(t_1) \) and \( X(t_2) \) are independent of the relative times \( t_1 \) and \( t_2 \). The independence of time is obtained by the dependence of the statistics on only the time difference between the two times, not on the precise values of each.

First type of stationarity: Strict sense stationary (SSS). This is the strongest. A RP is said to be strict sense stationary, if the statistical properties are invariant wrt time translation.
That is, $X(t,s)$ is SSS if for every value of $N$ and for every set of time instants $\{t_i \in T, i = 1, 2, \ldots, N\}$

$$F_{X(t_1), X(t_2), \ldots, X(t_N)}(x_1, x_2, \ldots, x_N) = F_{X(t_1+\tau), X(t_2+\tau), \ldots, X(t_N+\tau)}(x_1, x_2, \ldots, x_N)$$

for all values of $x_1, x_2, \ldots, x_N$ and for $\tau$ such that $t_j + \tau \in T$. This statement is simply a mathematical statement of the property which has already been stated, that the statistics depends only on the time differences. Such statistics are preserved if the time values are translated by the same amount.

SSS often unnecessarily restrictive, since most of the important results for real-world applications of RP based only on second-order terms, or terms involving only two time instants. A weaker definition involving only first and second order moments will be given shortly.

Note that proving SSS can be hard. Proving non-stationarity (NS) often easy.

Example: Consider sine wave process with random amplitude given as

$$X(t) = Y \cos(\omega_o t) \quad -\infty < t < \infty$$

Let $Y$ be uniformly distributed on $[0, 1]$. Easily seen that $E[X(t)] = 1/2 \cos \omega_o t$:

$$E[X(t)] = E[Y \cos(\omega_o t)] = \int_{-\infty}^{+\infty} y \cos(\omega_o t) f_y(y) dy = \int_0^1 y \cos(\omega_o t)(1) dy = \cos(\omega_o) t \int_0^1 y dy = \frac{1}{2} \cos(\omega_o) t$$

First moment function of time. Therefore NS.

Other example show $X(t) = \cos(\omega_o t + \theta), \theta$ RV. This is a SSS process.

Description of Random Processes

The complete set of joint probability density functions give the full description of a random process. But this are not always available. As a matter of fact, that set is rarely available. In practice, much of the study of random processes is based on second-order theory, for which only the joint statistics at two different instants of time are needed. These statistics are adequate for computing mean values of power. Furthermore, frequency spectra are also adequately described by these second-order statistics. Note that an important process that is completely described by second-order statistics is the Gaussian random process.

Autocorrelation

The term $E[X(u)X(v)]$ is a fundamental quantity in the consideration of RPs. It is called the autocorrelation function of the random process $X(t)$ and is defined, for time instants $t_1$ and $t_2$ as
\begin{align*}
R_x(t_1, t_2) &= E[X(t_1)X^*(t_2)] = \int_{-\infty}^{+\infty} x_1x_2f_{X(t_1)X(t_2)}(x_1, x_2)dx_1, dx_2
\end{align*}

Output process \(Y(t)\) of a linear system to an input \(X(t)\) is given by,
\begin{align*}
R_y(t_1, t_2) &= \int_{-\infty}^{+\infty} R_x(t_1 - u, t_2 - v)h(u)h(v)dudv
\end{align*}

For the WSS case, it can be easily shown that,
\begin{align*}
R_y(t) &= R_x(t) * h(t) * h(-t)
\end{align*}
from which we can find average output power \(R_y(0)\) of some linear system.

It is also easily seen that in terms of power spectral density \(S_x(f)\), the Fourier transform of \(R_x(\tau)\), and the transfer function of a filter \(H(f)\), we have
\begin{align*}
S_y(f) &= |H(f)|^2S_x(f)
\end{align*}
To compute mean-square output \(R_y(0) = E[Y^2(t)]\) of a linear system, need autocorrelation \(R_x(\tau)\) of input.

Covariance
\begin{align*}
K_x(t_1, t_2) &= \text{Cov}[X(t_1)X(t_2)] = E\{[X(t_1) - E[X(t_1)]][X(t_2) - E[X(t_2)]]\}
&= R_x(t_1, t_2) - E[X(t_1)]E[X(t_2)]
&= R_x(t_1, t_2) - m_x(t_1)m_x(t_2)
\end{align*}

SSS says that the density or probability mass function are invariant under time translation, implies that for SSS, autocorrelation function depends only on time difference between \(t_1, t_2\).
The converse not necessarily true. However, for Gaussian RP, the converse is true.

Wide-sense stationarity

RP \(X(t)\) is WSS if the mean function \(m_x(t) = E[X(t)]\) does not depend on \(t\) (i.e. \(m_x(t)\) is constant) and \(R_x(t_1, t_2) = R_x(\tau)\) is a function of \(\tau = t_2 - t_1\). So, if \(X(t)\) is WSS,
\begin{align*}
R_x(\tau) &= E[X(t)X(t + \tau)] \quad \forall \quad \tau.
\end{align*}

Some other properties and relationships

1. Sum of RVs
$$Z = X + Y \rightarrow f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(x, z - x)dx$$

2. For $X$ and $Y$ statistically independent,

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z - x)dx$$

3. Random variables $X$ and $Y$ said to be orthogonal if $E(XY) = 0$.
(Note that in using the term ‘orthogonal’ we are assuming a geometrical interpretation of RVs. That is, we are treating them as vectors in a vector space. Hence, the second moment $E[XY]$ is interpreted as an inner product, $E[X^2]$ as square of the length, etc.)

4. $\text{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\} = E(X_iX_j) - E(X_i)E(X_j)$

5. Random variables $X$ and $Y$ are said to be independent if $f_{XY}(xy) = g_X(x)h_Y(y)$ where $g_X(x)$ and $h_Y(y)$ are the pdfs of random variables $X$ and $Y$.

6. If $\text{Cov}(X_i, X_j) = 0$ then the random variables $X_i, X_j$ are said to be uncorrelated. For $X_1, X_2, \ldots, X_N$ uncorrelated, (weaker than independence), we have

$$\text{Var} \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} \text{Var}(X_i)$$

that is, variance of sum = sum of variances.

7. If $X, Y$ both orthogonal and uncorrelated, either $X$ or $Y$ must have zero mean.

8. If Gaussian variables $X$ and $Y$ are uncorrelated then they are also independent.

9. Correlation coefficient of random variables $X$ and $Y$ denoted by $\rho_{xy}$ or $\rho$, defined as

$$\rho = \rho_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Gaussian Processes
Let \( X \) be a \( N \times 1 \) column vector of random variables. The parameters of the \( N \)-dimensional Gaussian density function are the mean \( \mu_x^T = E(X^T) \) and \( \sum_x = E(XX^T) \) - the \( N \times N \) covariance matrix (matrix of variances and covariances). The \( N \)-dimensional Gaussian density function is written as

\[
   f_X(x) = \frac{1}{(2\pi)^{N/2} \sqrt{|\sum_x|}} \exp\left[\frac{-(x - \mu_x)^T \sum_x^{-1} (x - \mu_x)}{2}\right]
\]

So, for a Gaussian process, the complete statistical description of \( X(t_i), i = 1, 2, \ldots, N \) depends only on the mean vector \( \mu_x \) and covariance matrix \( \sum_x \). (Equivalently, the mean vector \( m_x(t) \) and covariance function \( C_x(t_1, t_2) \).)

If \( X(t) \) is WSS, then \( m_x = m, C_x(t_1, t_2) = C_x(\tau) \) where \( m_x \) is a constant vector and \( \tau = (t_2 - t_1) \).

Examples:

1. Simple case for the covariance matrix is when all the variables are uncorrelated and have variance 1, that is

   \[
   \sum_x = I
   \]

   Not hard to see that for this case, the joint Gaussian density function is

   \[
   f_X(x) = \prod_{i=1}^{N} f_{X_i}(x_i)
   \]

   Thus all the RVs are statistically independent and all have variance 1.

2. Two dimensional example of correlated Gaussian RVs \( X^T = (X_1, X_2) \), with \( \mu_x^T = (0, 0) \) and

   \[
   \sum_x = \begin{bmatrix} 5 & 6 \\ 6 & 9 \end{bmatrix}
   \]

   and the joint density function becomes
\[ f_X(x) = f_{X_1,X_2}(x_1,x_2) = \frac{1}{6\pi} \exp\left[ \frac{-\left(9x_1^2 - 12x_1x_2 + 5x_2^2\right)}{18} \right] \]

Of course, for a Gaussian random vector, the mean vector and covariance matrix convey the same information as the joint density function.

3. Another consequence of the definition of a Gaussian random process is that if \( X(t) \) is a GRP, the \( N \) random variables \( X(t_1), X(t_2), \ldots, X(t_N) \) are jointly Gaussian random variables. Frequently, this property is used as a definition of a Gaussian random process; i.e. a GRP is a process for which the \( N \) random variables \( X(t_1), X(t_2), \ldots, X(t_n) \) are jointly GRVs.

4. If \( X(t) \) is a WSS then \( m_x(t) = m_x \), a constant and \( R_x(\tau) \) is a function only of \( \tau = t_2 - t_1 \). Easily seen that the covariance \( K_x(\tau) \) also depends only on \( \tau \). Since these two functions also define the joint PDF, we see that \( \text{WSS} \rightarrow \text{SSS} \) for a GRP.

5. A random process said to be ergodic if time averages can be used to replace statistical averages (mean, autocorrelation, etc..). If so, then RP must be stationary. Thus if process is ergodic, then it is stationary. Converse not true. Stationary process does not mean that process has to be ergodic.

So, we calculate time averages, that is, the mean and autocorrelation function as follows:

\[ m_x = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{+T} X(t)dt \quad T \to \infty \]

and

\[ R_x(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{+T} X(t)X(t+\tau)dt \quad T \to \infty \]

Ergodicity useful since various quantities (mean, autocorrelation, etc.) can be measured from the actual sample functions. In fact, all real correlators are based on time averages and as such, most random processes in engineering are assumed to be ergodic.

6. Some properties of autocorrelation function of real WSS RP. (Usual modifications for complex WSS RP.)

\[ R_x(0) = E[(X^2(t))] = \int_{-\infty}^{\infty} x^2 f_{X(t)}(x)dx \]

\[ R_x(\tau) = R_x(-\tau) \]

\[ |R_x(\tau)| \leq R_x(0) \]

7. Spectral density

\[ R_x(\tau) \leftrightarrow S_x(f) \]

7
Some properties:

\[
S_x(f) \geq 0 \\
S_x(-f) = S_x(f) \\
R_x(0) = \int_{-\infty}^{+\infty} S_x(f) df = E[X^2(t)]
\]

The last property justifies the name power spectral density since the integral over the frequency yields power (mean square value).

8. White noise

A zero-mean noise RP is said to be white noise if its power spectral density is constant for all frequencies, i.e. if

\[
S_n(f) = \frac{N_o}{2} \quad -\infty < f < +\infty \\
R_n(\tau) = \frac{N_o}{2} \delta(\tau)
\]

where division by 2 is used when both negative and positive frequencies are considered. Use of white noise RP is similar to use of a \( \delta \) function in impulse response for linear systems. White noise is uncorrelated for all \( \tau \neq 0 \).

The most common assumption about the type of pdf associated with noise is that it is Gaussian.

(J.Komo - Random signal analysis in Eng. Systems, Ch 7,4)

Class notes:mirchand/ee276-2003