Random Processes

Jose C. Principe

Random Variable

Is a real function defined on the events of a probability space.

Cumulative Distribution Function

For each point a, it is the probability of the random variable X being less or equal than a, that is

$$\forall a, \mathbf{F}_X(a) = P(X \leq a) \tag{1}$$

Properties of $\mathbf{F}_X(x)$:

- always positive
- non-decreasing
- right-continuous

Density Function

Is a function defined on the random variable X, for which

$$f(x) \geq 0 \tag{2}$$

and

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{3}$$

If the r.v. X is discrete then these conditions are equivalent to $f(x_i) \geq 0$ and $\sum_i f(x_i) = 1$.

The density function is the derivative of $\mathbf{F}_X(x)$.

$$\begin{aligned} \mathbf{F}(a) &= \int -\infty^a f(x)dx \\ \frac{d\mathbf{F}_X(x)}{dx} &= f(x) \end{aligned} \tag{4}$$

or $\mathbf{F}_X(a) = \sum_{x_i \leq a} f(x_i)$ if X is discrete.

Expectation

Measure the probability of occurrence of an event

$$E[X] = \sum_i x_i p(x_i) \tag{5}$$

Expectation of a discrete function is:

$$E[h(X)] = \sum_i f(x_i) f(x_i) \tag{6}$$

Expectation of a continuous function is:

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx \tag{7}$$

Properties:

- $E[.]$ is a linear operator
- $E[h_1(X) + h_2(X)] = E[h_1(X)] + E[h_2(X)]$
- $E[Ch(X)] = CE[h(X)], C \in \mathbb{R}$
- $E[C] = C, C \in \mathbb{R}$
- $E[h(X)] \geq 0$, if $h(X) \geq 0$

Moments

How do you characterize an arbitrary density function? Answer: by using the r^{th} moment about the origin

$$\begin{aligned} u'_r &= E[x_i^n] = \sum_i x_i^n f(x_i) \\ u'_r &= E[x^n] = \int_{-\infty}^{\infty} x^n f(x)dx \end{aligned} \tag{8}$$

First moment is the **mean**

$$u = u'_1 = E[X] = \inf_{-\infty}^{\infty} xf(x)dx \tag{9}$$

The r^{th} moment around the means is

$$\begin{aligned} u_r &= E[(x_i - u)^r] = \sum_i (x_i - u)^r f(x_i) \\ u_r &= E[(X - u)^r] = \int_{-\infty}^{\infty} (x - u)^r f(x)dx \end{aligned} \tag{10}$$

Variance is the second moment around the mean

$$\sigma^2 = u_2 = E[(X - u)^2] = \int_{-\infty}^{\infty} (x - u)^2 f(x)dx \tag{11}$$

Bivariate Distribution

Measures statistical relations between two random variables X and Y. The **joint distribution** function $F_{XY}(x, y)$ is defined as

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \tag{12}$$

The **joint density** function $f_{XY}(x,y)$ is defined as

$$\begin{cases} f_{XY}(x,y) \geq 0 \\ \sum_i \sum_j f_{XY}(x_i, y_j) = 1 \end{cases} \tag{13}$$

$$\begin{cases} f_{XY}(x,y) \geq 0 \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y) dxdy \end{cases} \tag{14}$$

Expectation:

$$\begin{aligned} E[h(X,Y)] &= \sum_X \sum_Y h(x,y) f(x,y) \\ E[h(X,Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y) dxdy \end{aligned} \tag{15}$$

Bivariate expectation is still a linear operator.

Independence

Two r.v. are independent if

$$\begin{aligned} P(X \leq x, Y \leq y) &= P(X \leq x) P(Y \leq y) \\ &\Rightarrow \begin{cases} \mathbf{F}(X,Y) = \mathbf{F}(X)\mathbf{F}(Y) \\ f(X,Y) = f(X)f(Y) \end{cases} \end{aligned} \tag{16}$$

If independent, then $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$. The covariance (second mixed moment) is defined as

$$u_{11} = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \tag{17}$$

When X and Y are independent, the covariance is zero. The normalized covariance is called **correlation coefficient**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{18}$$

and $-1 \leq \rho \leq 1$.

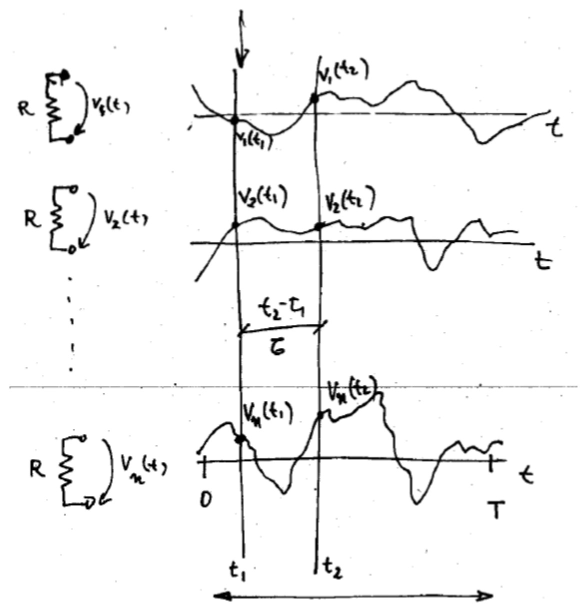
Random or Stochastic Process

Is a set of functions of some parameter (time), together with a probability measure. When the parameter is time, random processes are called **stochastic processes.**

Example: Mean square value thermal noise

$$\begin{aligned} \vartheta^2 &= 4kTR\nabla f(volts)^2 \\ k &= \text{ Boltzman constant} \\ T &- \text{ absolute temperature} \\ R &- \text{ resistance } \Omega \\ \nabla f &- \text{ measurement bandwidth} \end{aligned} \tag{19}$$

Gaussian



Each member of the set is a **realization**. At any time t, the realization $\vartheta(t)$ is a realization of a random variable $\vartheta(t)$. Therefore, a random process can also be defined as an indexed family of random variables $\{x(t), t \in T\}$. Depending upon the index set, a r.p. can be discrete parameter or continuous parameter. The random variable can also be discrete or continuous.

Stationarity

random process

A r.p. is **stationary** if its statistical properties do not change with time.

If the properties do change with time, then the r.p. is called **non-stationary**. Stationarity conditions can be given in terms of the moments.

We can define the density function of the r.v. $v(t_1)$, $f_{T_1}(v(t_1))$, which is called the **1st order amplitude** probability density function.

A process is stationary to order in $[0, T]$ if

$$f_{T_1}(v(t_1)) = f_{T_1}(v(t_1 + h)) \forall h, t_1 + h \in [0, T] \tag{20}$$

Therefore such processes have constant mean

$$\boxed{E[v(t_1)] = E[v(t_1 + h)] = u} \tag{21}$$

Now consider all joint distributions $f_{T_1T_2}(v(t_1), v(t_2))$

$$f_{T_1T_2}(v(t_1), v(t_2)) = P(v(t_1) \leq v(t_1) \leq v(t_1) + dv(t_1), v(t_2) \leq v(t_2) \leq v(t_2) + dv(t_2)) \tag{22}$$

We define a process stationary of order two in $[0, T]$ is

$$f_{T_1T_2}(v(t_1), v(t_2)) = f_{T_1T_2}(v(t_1 + h), v(t_2 + h)) \tag{23}$$

The expected value is

$$\boxed{E[v(t_1)v(t_2)] = E[v(t_1+h)v(t_2+h)]} \tag{24}$$

and it is called the **auto-correlation function**. For stationary processes of order two, the auto-correlation function only depends on the time lag $(t_1 - t_2)$, that is

$$\begin{aligned} R(t_1,t_2) &= E[v(t_1)v(t_2)] = R(t_1+h, t_2+h) \\ \text{For } h &= -t1R(t_1,t_2) = R(t_2 - t_1) \end{aligned} \tag{25}$$

When $E[v(t)]$ is zero, $R(t_1,t_2)$ is called the **covariance function.** One can define stationarity to order n in $[0,T]$ in the same way: $f_{T_1T_2\ldots T_n}(v(t_1)v(t_2)\ldots v(t_n)) = f_{T_1T_2\ldots T_n}(v(t_1+h)v(t_2+h)\ldots v(t_n+h))$.

A r.p. that is stationary to all orders is called **strictly stationary**. Normally, we are only interested in the case $n = 2$.

A r.p. is called **wide sense stationary** if

$$\begin{aligned} E[|v(t_i)^2|] &< \infty \forall t_i \\ E[v(t_1)v(t_2)] &= R(t_1 - t_1) \forall t_1, t_2 \end{aligned} \tag{26}$$

It is often convenient to also require $E[v(t)] = u = constant$.

Properties of Auto-Correlation Function

$$R(\tau) = E[v(t)v(t+\tau)], \text{ with } t_2 - t_1 = \tau \tag{27}$$

- The mean square value is $R(0)$

$$R(0) = E[v(t_1)^2] \tag{28}$$

- The auto-correlation function is even

$$R(\tau) = R(-\tau) \tag{29}$$

- The auto-correlation function is maximum at the origin

$$R(0) \geq |R(\tau)| \tag{30}$$

Example: Let $A4$ and θ be two r.v. A follows an uniform distribution over the interval $\{-A_0, A_0\}$ and θ follows an uniform distribution over the interval $\{-\pi, \pi\}$. Consider $e(t) = A\sin(wt+\theta)$.

Case I: θ is a r.v. and A is constant. Is $e(t)$ stationary?

$$\begin{aligned} R(t,\tau) &= R(t_1,t_2) \\ E[e(t_1)e(t_2)] &= E[e(t)e(t+\tau)] \\ &= E[A^2\sin(wt+\theta)\sin(w(t+\tau)+\theta)] \\ &= \frac{1}{2}E[A^2\cos(w\tau) - A^2\cos(2wt+2\theta+w\tau)], \text{ knowing that } \sin(x)\sin(y) = \frac{1}{2}\cos(x-y) - \frac{1}{2}\cos(x+y) \\ &= \frac{A^2}{2}\cos(w\tau) - \frac{A^2}{2}\int_{-\pi}^{\pi}\cos(2wt+2\theta+w\tau)\frac{1}{2\pi}d\theta \\ &= \frac{A^2}{2}\cos(w\tau) \end{aligned} \tag{31}$$

So, when θ is a r.v. and A is constant, $e(t)$ is wide sense stationary.

Case II: A is a r.v., w and θ are constants. Is $e(t)$ stationary?

$$R(t, t+\tau) = \frac{1}{2}\left[\cos(w\tau) - \cos(2wt + 2\theta + w\tau)\right] \int_{-A_0}^{A_0} A^2 \frac{1}{2A_0} dA \tag{32}$$

It depends on t, so it is non-stationary.

Ergodicity

if random process not stationary, not ergodic

To obtain any of the previous measures one needs statistical averages, i.e. the whole r.p.. This may not be feasible.

The assumption of ergodicity allows us to replace statistical averages by time averages in one realization of the r.p..

A r.p. is ergodic if the time averages in any of the realizations are equal to the statistical averages.

Ergodicity implies stationarity.

In engineering we almost always assume ergodicity.

Time Averages

We define time average of r.p. $v(t)$ as

$$\begin{aligned} A\{v(t)\} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t)dt \\ A\{v_n\} &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} v_n \end{aligned} \tag{33}$$

The time auto-correlation functions is

$$\begin{aligned} R(\tau) &= A\{v(t)v(t+\tau)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t)v(t+\tau)dt \\ R(\tau) &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} v_n v_{n+m} \end{aligned} \tag{34}$$

and the time cross-correlation function

$$R_{uv}(\tau) = A\{u(t)v(t)\} \tag{35}$$

For ergodic wide sense stationary r.p.

$$\begin{aligned} E[v(t)v(t+\tau)] &= A\{v(t)v(t+\tau)\} \\ E[v(t)] &= A\{v(t)\} = u \end{aligned} \tag{36}$$

Cross-Correlation Function

Two r.p. $u(t)$ and $v(t)$ with auto-correlation functions $R_{uu}(t_1, t_2)$ and $R_{vv}(t_1, t_2)$. The cross-correlation functions is

$$R_{uv}(t_1, t_2) = E[u(t_1)v(t_2)] \tag{37}$$

or

$$R_{vu}(t_1, t_2) = E[v(t_1)u(t_2)] \tag{38}$$

Properties:

- $R_{uv}(\tau) = r_{vu}(-\tau)$
- $|R_{uv}(\tau)| \leq \frac{1}{2}[R_{uu}(0) + R_{vv}(0)]$
- $|R_{uv}(\tau)|^2 \leq R_{uu}(0)R_{vv}(0)$

Fourier Transforms

If one takes the Fourier transform of the r.p. $x(t)$ we get

$$X(w) = \int_{-\infty}^{\infty} x(t)e^{-jwt}dt \tag{39}$$

$X(w)$ is another r.p., now complex, with real parameter w.

If the r.p. $x(t)$ is at least wide sense stationary with constant mean zero, we can define power spectral density as the Fourier transform of the auto-correlation function.

$$\begin{aligned} \varphi(w) &= \int_{-\infty}^{\infty} R(\tau)e^{-jw\tau}d\tau \\ R(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(w)e^{jw\tau}d\tau \end{aligned} \tag{40}$$

We can also show that the same result is valid for the time auto-correlation function

$$\begin{aligned} \varphi(w) &= \int_{-\infty}^{\infty} R(\tau)e^{-jw\tau}d\tau \\ R(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(w)e^{jw\tau}d\tau \end{aligned} \tag{41}$$

and also that

$$\begin{aligned} \varphi(w) &= \lim_{T\to\infty} E\left[\frac{|X_T(w)|^2}{2T}\right] \\ \text{where } X_T(w) &= \int_{-T}^{T} x(t)e^{-jwt}dt \end{aligned} \tag{42}$$

Spectrum of Discrete Time Infinite Energy Signals

Recall that the mean is defined as

$$
\begin{gathered}
E[v_n] = u \\
A\{v_n\} = \lim_{N \to 0} \frac{1}{2N+1} \sum_{n=-N}^{N} v_n
\end{gathered} \tag{43}
$$

and the auto-correlation function

$$
\begin{aligned}
R(m) &= E[v_n v_{n+m}] \\
C(m) &= E[(v_n - u)(v_{n+m} - u)] = R(m) - u^2
\end{aligned} \tag{44}
$$

Let us define the z transform of the auto-covariance function (or the auto-correlation function for zero mean r.p.)

$$
\begin{cases}
\gamma_{vv}(z) = \sum_{n=-\infty}^{\infty} C(n) z^{-n} \\
C(n) = \frac{1}{2\pi j} \phi_c \gamma_{vv}(z) z^{n-1} dz
\end{cases} \tag{45}
$$

Properties:

1. The integral of the power spectral density is the average power of the r.p..

$$
R(0) = E[x^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(w) dw \tag{46}
$$

2. The power spectral density is real.

$$
\varphi(w) = 2 \int_{0}^{\infty} R(\tau) \cos(w\tau) d\tau \tag{47}
$$

3. The power spectral density is non negative.
4. The power spectral density is an even function.

This function has nice properties

- $\sigma_v^2 = \frac{1}{2\pi j} \phi_c \gamma_{vv}(z) z^{-1} dz$
- $\gamma(z) = \gamma\left(\frac{1}{z^*}\right)$
- $\gamma_{uv}(z) = \gamma_{vu}\left(\frac{1}{z^*}\right)$

For discrete time signals, the power density spectrum is defined as the z-transform of the auto-covariance function calculated on the unit circle

$$
\gamma(e^{jw}) = \gamma(z)|_{z=e^{zw}} = \sum_{n=-\infty}^{\infty} C(n) e^{-jwn} \tag{48}
$$

can integrate in w, i.e.

$$
P(w) = \gamma(e^{jw}) \Rightarrow \sigma_v^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(w) dw \tag{49}
$$

The area under $P(w)$ is proportional to the average power.

- $P(w) = P(-w)$
- $P(w)$ is non negative

Response of Linear Systems to Random Signals

Consider a shift invariance linear system with constant coefficients

$$y(n) = \sum_{k=-\infty}^{\infty} h(n-k)x(k) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \tag{50}$$

If the input is stationary, then the output is also stationary. In fact, the output mean is

$$m_Y = E[y(n)] = \sum_k h(k)E[x(n-k)] = m_X \sum_k h(k) = m_X H(e^{j\theta}) \tag{51}$$

and the output auto-correlation function is

$$\begin{aligned} R_Y(n, n+m) = E[y(n)y(n+m)] = E[\sum_k \sum_r h(k)h(n)x(n-k)x(n+m-r] \\ E[x(n-k)x(n+m-k)] \to R(m+k-r) \Rightarrow \sum_k h(k) \sum_r h(r)R(m+n-r)) \end{aligned} \tag{52}$$

In the frequency domain, this result is much more readily interpreted.

$$\begin{aligned} \gamma_{yy}(z) &= V(z)\gamma_{xx}(z) \\ V(z) &= \mathcal{Z}\left[\sum_k h(k)h(k+l)\right] = H(z)H(z^{-1}) \\ \text{or } P_{yy}(w) &= |H(e^{jw})|^2 P_{xx}(w) \end{aligned} \tag{53}$$

Estimation

The problem is that the time auto-correlation function (or mean) can not be computed accurately (infinite time). **What is the error that we make when we use finite time to compute the mean and time auto-correlation function?** This problem is studied in spectral analysis, more particularly in estimation theory. Here just is a brief review.

If we use

$$\hat{m}_X = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \text{ (maximum likelihood)} \tag{54}$$

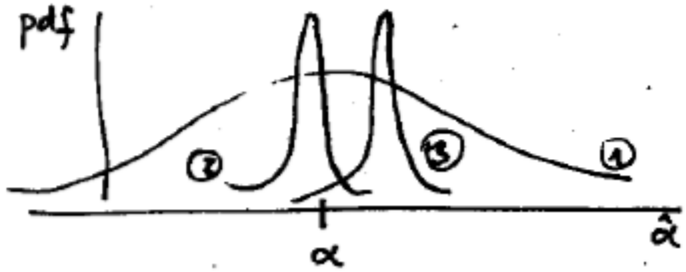
to compute the time mean, if N is sufficiently large, this is a good estimate. However, $\hat{m}_X$ is a r.v.. So, we can define its mean and variance (or its pdf). Different values of computing the mean will have different pdfs. From the following three examples, which one would we want?

So quality factors are the **bias** and the **variance** respectively defined as

$$\begin{aligned} B &= bias = \alpha - E[\hat{\alpha}] \\ \sigma_{\hat{\alpha}}^2 &= E[(\hat{\alpha} - E[\hat{\alpha}])^2] \end{aligned} \tag{55}$$

(an unbiased estimator has $B = 0$).

If the bias and the variance go to zero as the number of observations increases, then the estimator is called **consistent**. Maximum likelihood estimators are normally used.



Let us define the windowed auto-correlation function of $x(n)$ as

$$C_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n)x(n+m) \tag{56}$$

This function is a **biased estimated** of the true (infinite) auto-correlation function

$$E[C_{xx}(m)] = \frac{N-|m|}{N} R_{xx}(m) \Rightarrow B = R_{xX}(m) \left[\frac{m}{N}\right] \tag{57}$$

and the variance of the estimate also depends on N $\left(\propto \frac{1}{N}\right)$. However, for large N, the mean and the variance approach the true values of the auto-correlation function. So the windowed auto-correlation function is a consistent estimator of the auto-correlation function.

We could think that the Fourier transform of the windowed auto-correlation function was also an asymptotically unbiased estimator of the power spectral density. Unfortunately, this is not the case.

The Fourier transform of $C_{xx}(n)$ is called the **periodogram**

$$I_N(w) = \sum_{m=-(N-1)}^{N-1} C_{xx}(m) e^{-jNm} \tag{58}$$

It turns out that it can also be computed as

$$\begin{aligned} E_N(w) &= \frac{1}{N} |X(e^{jw})|^2 \\ X(e^{jw}) &= \sum_{n=0}^{N-1} x(n) e^{-jwn} \end{aligned} \tag{59}$$

This formula tells us that one way to efficiently compute the periodogram is with the FFT (just square it and take the absolute value).

However the periodogram is not a consistent estimator of the power spectrum (i.e. it may show peaks that do not correspond to resonances in the real data). However there are ways to decrease the variance through smoothing (Welch, etc.)