Chapter 4

Stochastic Processes

Clearly data assimilation schemes such as Optimal Interpolation are crucially dependent on the estimates of background and observation error statistics. Yet, we don't know what these covariances are, nor do we have enough observations to be able to estimate them. Thus we are left to parameterize or approximate them using some simplifying assumptions such as stationarity, or homogeneity or isotropy. These assumptions may not be valid, but some sort of simplification is necessary in order to reduce the number of unknown parameters to a number small enough to be estimated using the observations.

As we saw in the chapter on OI, we modelled the background error covariances by defining a continuous autocorrelation function which could be evaluated at model grid points or observation locations to produce correlation matrices. Thus, it would be useful to know more about autocorrelation functions for continuous processes. This subject comes from time series analysis or signal processing. The parameter is then time, rather than spatial dimensions. Since time is only 1 parameter and space can be up to 3 parameters, it is definitely easier to learn about covariance functions in the time series analysis context. At the end of this chapter, we briefly note the parallels for random fields (another entire subject unto itself) without going into details.

4.1 Definitions

A time varying random variable $\mathbf{x}(t)$ is called a **random process**. That is, at each instant in time, $\mathbf{x}_t$ is a random variable with no precise value, but rather a range of possible values, each with an associated p.d.f. describing the likelihood of each possible value. Random processes arise because

1. The physical system generating the process may have some random elements, e.g. the Brownian motion of particles suspended in a liquid.
2. The observation of the system involves measurement error.
3. The system may be deterministic but too complex to be completely specified.

More precisely, the random process $\mathbf{x}(t, \omega)$ is defined on a product space T x Ω where T is a subset of $\mathcal{R}$ and Ω is a sample space. However, for simplicity, we will drop the explicit reference to the events and speak of the random process, $\mathbf{x}(t)$.

Suppose we observe $\mathbf{x}(t)$ from an experiment that can be repeated under identical conditions. Then $\boldsymbol{x}(t)$ is only 1 possible realization of many possible realizations which together form an ensemble (see Fig. 4.1).

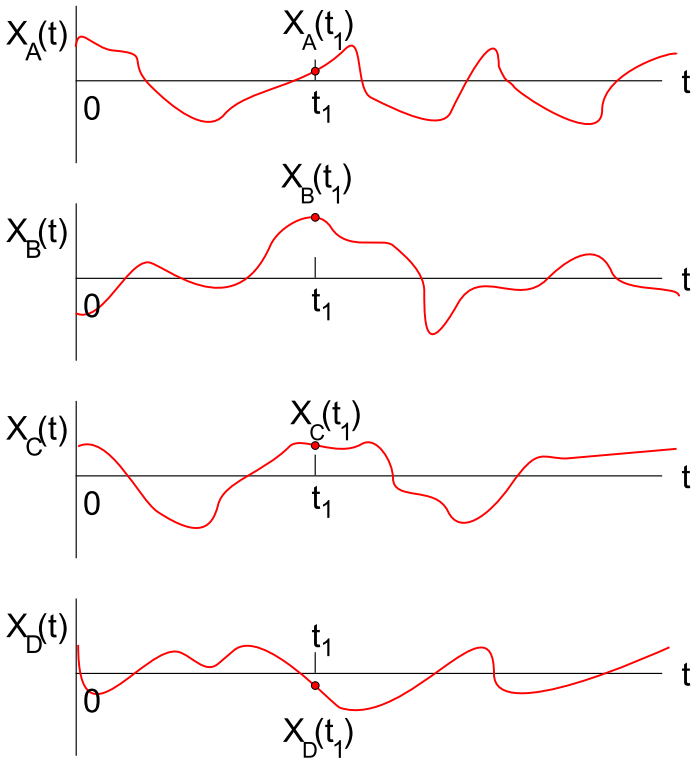

Figure 4.1: Ensemble of sample realizations of a random process. (After Brown and Hwang (1977) Fig. 2.2, pg. 74)

Example 4.1 *Card Game*

A player takes a card from a deck, records the value (1-13) and returns it to the deck. The deck is reshuffled and the experiment repeated. This defines a random process which is discrete in time and amplitude.

Example 4.2 *Signal*

Consider a signal generated as follows:

1. *There is a sample and hold pattern with a hold of 1 second.*
2. *The amplitude is random from a uniform distribution between [-1,1].*
3. *The first switching time after $t = 0$ is a random variable with a uniform distribution from 0 to 1.*

A sample realization of such a process is depicted in Brown and Hwang (1997)'s Fig. 2.3.

The mean of the process is

$$\mu_x(t) = \int_{-1}^{1} x(t) p_{\mathrm{x}(t)}(x) dx = \int_{-1}^{1} \frac{1}{2} x(t) dx = \left. \frac{1}{4} x^2 \right|_{-1}^{1} = 0. \tag{4.1}$$

Since the mean is 0, the variance is given by E(x^2).

$$\sigma_x^2 = \int_{-1}^{1} x^2(t) p_{\mathrm{x}(t)}(x) dx = \int_{-1}^{1} \frac{1}{2} x^2(t) dx = \left. \frac{1}{6} x^3 \right|_{-1}^{1} = \frac{1}{3}. \tag{4.2}$$

The variance of the process at any instant, t is 1/3.

Example 4.3 *Another signal*

Consider a similar signal except

1. *The hold pattern lasts 0.2 seconds,*
2. *The amplitude is random from a Gaussian distribution with mean 0 and variance 1/3.*
3. *The first switching time after $t = 0$ is not random but at 0.2 seconds.*

A sample realization of this process is also found in Brown and Hwang (1997)'s Fig. 2.4.

The mean and variance of the Gaussian at any time t are 0 and 1/3.

In the last two examples, both processes had the same mean and variance yet they were clearly very different processes. The first one varies much more slowly in time than the second one. Clearly it is insufficient to describe a process by a mean and variance alone. What else is needed? It would be useful to know how quickly values change with time for a given realization, i.e.

$$F_2(\mathrm{x}_1, t_1; \mathrm{x}_2, t_2) = P[\mathrm{x}(t_1) \leq x_1, \mathrm{x}(t_2) \leq x_2]$$
$$p_2(\mathrm{x}_1, t_1; \mathrm{x}_2, t_2) = \frac{\partial^2 F_2(\mathrm{x}_1, t_1; \mathrm{x}_2, t_2)}{\partial x_1 \partial x_2}.$$

Here p_2 is a second-order joint p.d.f. Higher order p.d.f.'s can be similarly defined but rarely does one attempt to deal with anything higher than second order.

Since it is difficult to completely define a p.d.f. let alone higher order p.d.f.'s, we can at least try to indicate how the process is correlated with itself at different times. This is what the autocorrelation function describes. Fig. 4.4 of Maybeck illustrates the autocorrelation functions for two different processes. When $t_1 = t_2$, the autocorrelation function depicts the variance of the process. If the autocorrelation function drops off quickly with $(t_1 - t_2)$, then the correlation between the process at two different times drops rapidly as the time interval increases. However, for a slowly decreasing autocorrelation function (see left panel of Maybeck's Fig. 4.4), the correlation between the process at different times drops slowly as the time interval increases.

4.2 The Auto-correlation Function

The auto-correlation function is the first moment or mean of the second order joint p.d.f. and is defined by

$$\mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) = E(\mathbf{x}(t_1)\mathbf{x}(t_2)^{\mathrm{T}}) = E(\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}}). \tag{4.3}$$

if we denote $\mathbf{x}_k = \mathbf{x}(t_k)$. In the scalar case, we could then write

$$\Gamma_{\mathrm{x}}(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_{\mathrm{x_1 x_2}}(x_1, x_2)dx_1 dx_2. \tag{4.4}$$

However, this is not always an easy way to evaluate the autocorrelation function because it requires knowledge of the p.d.f. Similarly, we could define a **cross-correlation** function between two random variables, $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{\Gamma}_{\mathbf{xy}}(t_1, t_2) = E(\mathbf{x}(t_1)\mathbf{y}(t_2)^{\mathrm{T}}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_{\mathrm{xy}}(x(t_1), y(t_2))dxdy. \tag{4.5}$$

where the second equality is for the scalar case. If E(x(t_1))=E(x(t_2))=E(y(t_2))=0, these correlation functions are the covariances of the random variables. If they are normalized by standard deviations, i.e.

$$\rho = \frac{E(\mathrm{xy}) - E(\mathrm{x})E(\mathrm{y})}{\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}}$$

they become correlation coefficients which vary from -1 to 1 and measure the linear dependence between variables.

In the case of vector-valued random processes (for example wind measurements, since winds can be viewed as (u,v,w)) the autocorrelation function is called the autocorrelation matrix. It is defined by

$$\begin{aligned} \mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) &= E(\mathbf{x}(t_1)\mathbf{x}(t_2)^{\mathrm{T}}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{z}\boldsymbol{y}^{\mathrm{T}} p_{\mathbf{x}(t_1)\mathbf{x}(t_2)}(\boldsymbol{z}, \boldsymbol{y})d\boldsymbol{z}d\boldsymbol{y} \end{aligned} \tag{4.6}$$

or

$$\begin{aligned} \mathbf{\Gamma}_{\mathbf{x}}(t, t+\tau) &= E(\mathbf{x}(t)\mathbf{x}(t+\tau)^{\mathrm{T}}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{z}\boldsymbol{y}^{\mathrm{T}} p_{\mathbf{x}(t)\mathbf{x}(t+\tau)}(\boldsymbol{z}, \boldsymbol{y})d\boldsymbol{z}d\boldsymbol{y}. \end{aligned} \tag{4.7}$$

The autocorrelation is the ensemble average of the product of $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$. Since we usually don't know the joint p.d.f. it is difficult to compute using the above. However, if the *ergodic* hypothesis is valid, we can replace ensemble averages with time averages.

4.2.1 Auto-correlation function properties

1. $\mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) = E\{\mathbf{x}(t_1)\mathbf{x}(t_2)^{\mathrm{T}}\} = [E\{\mathbf{x}(t_2)\mathbf{x}(t_1)^{\mathrm{T}}\}]^{\mathrm{T}} = \mathbf{\Gamma}_{\mathbf{x}}(t_2, t_1)^{\mathrm{T}}$.
2. $\mathbf{\Gamma}_{\mathbf{x}}(t, t) = E(\mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}}) = E(|\mathbf{x}(t)|^2) \geq 0$.
3. $\mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2)$ is positive semi-definite.

Proof of property (3):
Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$. For any $\mathbf{a}$ we have that

$$E\{|\sum_{i=1}^{n} a_i \mathbf{x}(t_i)|^2\} \geq 0$$

but

$$\begin{aligned} E\{|\sum_{i=1}^{n} a_i \mathbf{x}(t_i)|^2\} &= E\{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbf{x}(t_i)\mathbf{x}(t_j)^{\mathrm{T}} a_j\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i E\{\mathbf{x}(t_i)\mathbf{x}(t_j)^{\mathrm{T}}\} a_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) a_j \end{aligned}$$

Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) a_j = \mathbf{a}^{\mathrm{T}} \mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2)\mathbf{a} \geq 0..$$

In other words, $\mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2)$ is positive semi-definite.

4.3 Other definitions

The auto-covariance matrix is

$$\mathbf{C}_{\mathbf{x}}(t_1, t_2) = E[(\mathbf{x}(t_1) - \boldsymbol{\mu}_{\mathbf{x}}(t_1))(\mathbf{x}(t_2) - \boldsymbol{\mu}_{\mathbf{x}}(t_2)^{\mathrm{T}}] = \mathbf{\Gamma}_{\mathbf{x}}(t_1, t_2) - \boldsymbol{\mu}_{\mathbf{x}}(t_1)\boldsymbol{\mu}_{\mathbf{x}}(t_2)^{\mathrm{T}}.$$

Two random processes, $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are called **uncorrelated** if their cross-covariance is 0, i.e.

$$E[(\mathbf{x}(t_1) - \boldsymbol{\mu}_{\mathbf{x}}(t_1))(\mathbf{y}(t_2) - \boldsymbol{\mu}_{\mathbf{y}}(t_2))^{\mathrm{T}}] = 0.$$

The processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are **orthogonal** if the correlation matrix is 0,

$$E[\mathbf{x}(t_1)\mathbf{y}(t_2)^{\mathrm{T}}] = 0.$$

A process $\mathbf{x}(t)$ is **serially uncorrelated** if

$$E[(\mathbf{x}(t_1) - \boldsymbol{\mu}_{\mathbf{x}}(t_1))(\mathbf{x}(t_2) - \boldsymbol{\mu}_{\mathbf{x}}(t_2)^{\mathrm{T}}] = Q(t_1, t_2)\delta(t_1 - t_2)$$

where

$$\int_a^b \delta(t)dt = \begin{cases} 1 & \text{if } a \leq t \leq b \\ 0 & \text{otherwise.} \end{cases}$$

For discrete time processes a similar definition exists. A random sequence $\mathbf{x}(t)$ is **serially uncorrelated** if

$$E[(\mathbf{x}(k) - \boldsymbol{\mu}_{\mathbf{x}}(k))(\mathbf{x}(j) - \boldsymbol{\mu}_{\mathbf{x}}(j)^{\mathrm{T}}] = Q(k,j)\Delta(k-j)$$

where

$$\Delta(k-j) = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise.} \end{cases}$$

A stochastic process $\mathrm{x}(t)$ is **independent** when for all τ,

$$p_{\mathbf{x}(t)\mathbf{x}(\tau)}(\mathbf{x}(t), \mathbf{x}(\tau)) = p_{\mathbf{x}(t)} p_{\mathbf{x}(\tau)}.$$

i.e. for an $x_1, \ldots, x_n$,

$$p_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_{x(t_1)}(x_i).$$

If the process is independent, then it is uncorrelated in time. The reverse is NOT true however.

4.4 Stationarity

A process is stationary if the density functions are invariant in time. Consider

$$\mathbf{x}_1 = \mathbf{x}(t_1), \mathbf{x}_2 = \mathbf{x}(t_2), \ldots, \mathbf{x}_k = \mathbf{x}(t_k).$$

Translate this in time by τ:

$$\mathbf{x}'_1 = \mathbf{x}(t_1 + \tau), \mathbf{x}'_2 = \mathbf{x}(t_2 + \tau), \ldots, \mathbf{x}'_k = \mathbf{x}(t_k + \tau).$$

If the process is stationary, then the densities

$$p_{\mathbf{x}_1}(\mathbf{x}_1), p_{\mathbf{x}_1\mathbf{x}_2}(\mathbf{x}_1, \mathbf{x}_2), \ldots, p_{\mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_k}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$$

are the same for both time series. Thus, not only the p.d.f.s are the same, but also all higher order joint p.d.f.s are the same despite the translation in time. This is stationariy in the strict sense. However, this is very difficult to verify in practice, since it is difficult to determine higher order p.d.f.'s. In practice, we use **wide sense stationarity** (WSS) in which (1) the first moment is independent of time, and (2) the autocorrelation depends only on the time interval between events. Thus, for WSS stationary processes, the mean ($E[\mathbf{x}(t)] = \boldsymbol{\mu}$) is independent of time, and the auto-correlation is a function of time interval:

$$\boldsymbol{\Gamma}_{\mathrm{x}}(t, t+\tau) = \boldsymbol{\Gamma}_{\mathrm{x}}(\tau).$$

For a scalar process, $\mathrm{x}(t)$, we can define the autocovariance function as:

$$C_{\mathrm{x}}(\tau) = E[(\mathrm{x}(t) - \mu)(\mathrm{x}(t+\tau) - \mu)].$$

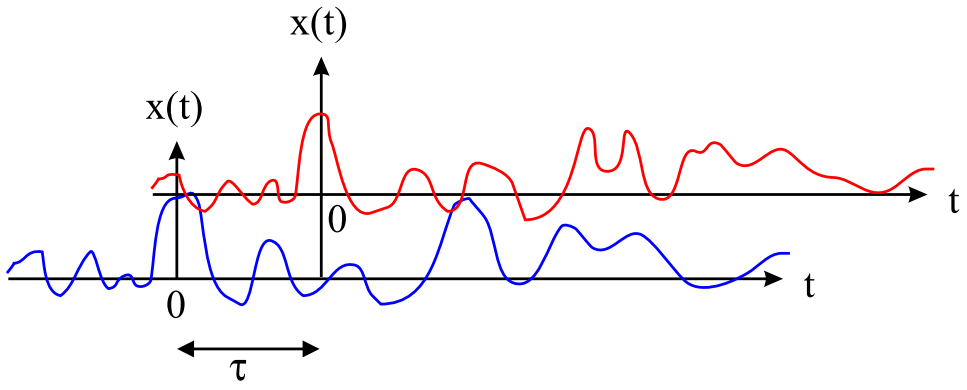

Figure 4.2: The autocovariance of a stationary process refers to the covariance of a process with itself, translated in time.

Fig. 4.2 illustrates the idea that the autocovariance function measures the covariance of a process with itself, translated in time.

Note that:

$$C_{\mathrm{x}}(0) = E[(\mathrm{x}(t) - \mu)^2] = \sigma^2$$

$$\rho_{\mathrm{x}}(\tau) \quad = \quad \frac{E[(\mathrm{x}(t) - \mu)(\mathrm{x}(t+\tau) - \mu)]}{\sqrt{E[(\mathrm{x}(t) - \mu)^2]}\sqrt{E[(\mathrm{x}(t+\tau) - \mu)^2]}} = \frac{C_{\mathrm{x}}(\tau)}{\sigma^2} = \frac{C_{\mathrm{x}}(\tau)}{C_{\mathrm{x}}(0)} \leq 1$$

Thus,

$$C_{\mathrm{x}}(\tau) \leq C_{\mathrm{x}}(0).$$

Fig. 4.3 depicts a generic correlation function for a continuous parameter process.

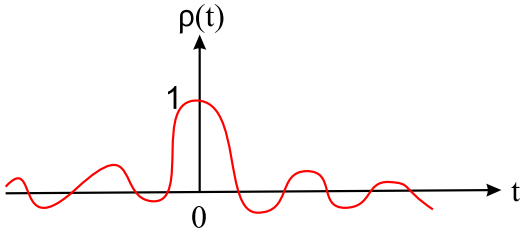

Figure 4.3: The autocorrelation function of a continuous parameter process. (After Priestley 1981, Fig. 3.4)

If $\mathbf{x}(t)$ is stationary, its auto-correlation function is given by

$$\mathbf{\Gamma}_{\mathbf{x}}(\tau) = E[\mathbf{x}(t)\mathbf{x}(t+\tau)^{\mathrm{T}}] = \mathbf{C}_{\mathbf{x}}(\tau) + \boldsymbol{\mu}_{\mathbf{x}}(t)\boldsymbol{\mu}_{\mathbf{x}}(t+\tau)^{\mathrm{T}}.$$

The auto-correlation function for a scalar stationary process, $\mathrm{x}(t)$ has the following properties.

1. $\Gamma_{\mathrm{x}}(0)$ is the mean square value of the process.

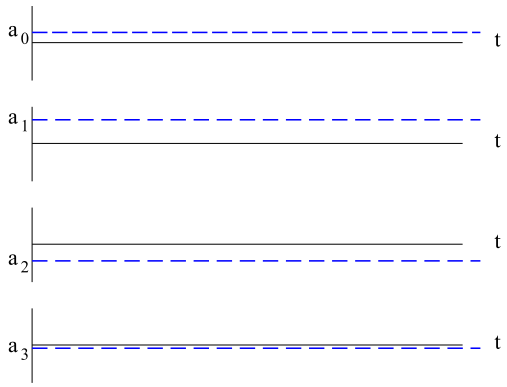

Figure 4.4: Some sample realizations of the process x(t)=c, where c is a random constant.

2. $\Gamma_{\mathrm{x}}(\tau) = \Gamma_{\mathrm{x}}(-\tau)$. Γ_{x} is an even function of τ.

3. $|\Gamma_{\mathrm{x}}(\tau)| \leq \Gamma_{\mathrm{x}}(0)$. (Since $|\rho_{\mathrm{x}}(\tau)| = |\frac{C_{\mathrm{x}}(\tau)}{C_{\mathrm{x}}(0)}| = |\frac{\Gamma_{\mathrm{x}}(\tau)+\mu^2}{\Gamma_{\mathrm{x}}(0)+\mu^2}| \leq 1$.)

4. If x(t) has periodic components, so does $\Gamma_{\mathrm{x}}(\tau)$. (Write x($t$) as a sum of periodic and nonperiodic components and compute $\Gamma_{\mathrm{x}}(\tau)$. A constant is a special type of periodic function.)

5. If x(t) has no periodic components, $\Gamma_{\mathrm{x}}(\tau) \to 0$ as $\tau \to \infty$. Thus, x(t), x($t+\tau$) become uncorrelated as $\tau \to \infty$ and the process mean is 0.

6. $\Gamma_{\mathrm{x}}(\tau)$ is positive semi-definite.

7. The Fourier transform of $\Gamma_{\mathrm{x}}(\tau)$ is real, symmetric and non-negative.

4.5 Ergodicity

A random process is **ergodic** if time averaging is equivalent to ensemble averaging. This means that a single realization contains all possible statistical variations of the process. In other words, there is no advantage in observing an ensemble of realizations.

Under the ergodic hypothesis,

$$\begin{aligned} E(\mathrm{x}(t)) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt \\ E(\mathrm{x}^2(t)) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)dt \\ \Gamma_{\mathrm{x}}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)dt \\ \Gamma_{\mathrm{xy}}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y(t+\tau)dt. \end{aligned}$$

Example 4.4 *Constant*

Consider the process, $\mathrm{x}(t) = c$*, where* c *is* $\mathcal{N}(0, \sigma^2)$*. This model is appropriate for random instrument bias. Consider some realizations of this process as depicted in Fig. 4.4. The time means are* a_0*,* a_1*,* a_2*,* a_3*. The ensemble mean is 0 (since the mean of* c *is 0). The process is NOT ergodic because time means and ensemble means are not equivalent. However, the process is stationary because its "statistics" do not change with time.*

Example 4.5 *(Gelb p.38)*

Consider

$$\mathrm{x}(t) = A\sin(\omega t + \theta)$$

where θ *is uniformly distributed in* $[0, 2\pi]$*. An ensemble average would find all* θ *represented with equal probability. Also a single realization of* $\mathrm{x}(t)$ *would find all* θ *equally represented. However if the distribution of* θ *were anything other than uniform, this would not be an ergodic process.*

The joint p.d.f. of $\mathrm{x}_1 = \mathrm{x}(t)$ *and* $\mathrm{x}_2 = \mathrm{x}(t+\tau)$ *is*

$$p(x_1, x_2) = \frac{1}{2\pi}, \qquad 0 \le \theta < 2\pi.$$

The ensemble average autocorrelation is:

$$\begin{aligned}
\Gamma_{\mathrm{x}}(t, t+\tau) &= \int_0^{2\pi} x(t)x(t+\tau)p(x(t), x(t+\tau))d\theta \\
&= \int_0^{2\pi} A\sin(\omega t + \theta)A\sin(\omega t + \omega\tau + \theta)\frac{1}{2\pi}d\theta \\
&= \frac{A^2}{4\pi}\int_0^{2\pi}[\cos(\omega\tau) - \cos(2\omega t + \omega\tau + 2\theta)]d\theta \\
&= \frac{A^2}{4\pi}2\pi\cos(\omega\tau) = \frac{A^2}{2}\cos(\omega\tau). \qquad (4.8)
\end{aligned}$$

To get the third line, we used the trigonometric equality

$$2\sin(a)\sin(b) = \cos(a-b) - \cos(a+b).$$

The time-averaged auto-correlation is

$$\begin{aligned}
\Gamma_{\mathrm{x}}(\tau) &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)x(t+\tau)dt \\
&= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A\sin(\omega t + \theta)A\sin(\omega t + \omega\tau + \theta)dt \\
&= \lim_{T\to\infty}\frac{A^2}{2T}\frac{1}{2}\int_{-T}^{T}[\cos(\omega\tau) - \cos(2\omega t + \omega\tau + 2\theta)]dt \\
&= \lim_{T\to\infty}\frac{A^2}{4T}2T\cos(\omega\tau) = \frac{A^2}{2}\cos(\omega\tau). \qquad (4.9)
\end{aligned}$$

The two results are the same so $\mathrm{x}(t)$ *is an ergodic process.*

Example 4.6 *Consider the process* $\mathrm{x}(t) = A\sin(\omega t)$ *where* A *is* $\mathcal{N}(0, \sigma^2)$*. Is this process ergodic?*

The ensemble average auto-correlation is

$$\begin{aligned} E(\mathrm{x}(t_1), \mathrm{x}(t_2)) &= E(A\sin(\omega t_1) A\sin(\omega t_2) \\ &= E(A^2)\sin(\omega t_1)\sin(\omega t_2) \\ &= \sigma^2 \sin(\omega t_1)\sin(\omega t_2). \end{aligned} \tag{4.10}$$

The time average auto-correlation is

$$\begin{aligned} \Gamma_{\mathrm{x}}(\tau) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} A\sin(\omega t) A\sin(\omega t+\tau) dt \\ &= \lim_{T\to\infty} \frac{A^2}{4T}\int_{-T}^{T} [\cos(\omega\tau) - \cos(2\omega t + \omega\tau)] dt \\ &= \lim_{T\to\infty} \frac{A^2}{4T} 2T\cos(\omega\tau) = \frac{A^2}{2}\cos(\omega\tau). \end{aligned} \tag{4.11}$$

The process is not ergodic since the time and ensemble average auto-correlation functions are not the same. Moreover, the ensemble average auto-correlation does not reduce to a function of $t_1 - t_2$ so the process is not stationary either.

4.6 The Wiener-Khintchine Relation

The power spectral density function is the Fourier transform of the auto-covariance function. For a stationary process:

$$\begin{aligned} S_{\mathrm{x}}(\omega) &= \int_{-\infty}^{\infty} C_{\mathrm{x}}(\tau) e^{-i\omega\tau} d\tau \\ C_{\mathrm{x}}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathrm{x}}(\omega) e^{i\omega\tau} d\omega. \end{aligned} \tag{4.12}$$

Because $C_{\mathrm{x}}(\tau)$ is a real and even function of τ (that is $C_{\mathrm{x}}(-\tau) = C_{\mathrm{x}}(\tau)$), $S_{\mathrm{x}}(\omega)$ is real and symmetric (or even) function of ω. It is also useful to note that

$$C_{\mathrm{x}}(0) = E[\mathrm{x}^2(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathrm{x}}(\omega) d\omega \geq 0.$$

Thus the area under the power spectrum of any process is positive. One can go further and show that the spectrum, $S_{\mathrm{x}}(\omega) \geq 0$ (e.g. see Papoulis 1965, p. 328). Moreover, a function, $S_{\mathrm{x}}(\omega)$ is a power spectrum if and only if it is positive.

One interpretation of the power spectrum is as an average of Fourier transforms of an ensemble of realizations. To see this, consider an ensemble of functions, $\mathrm{x}(t)$, over a finite time span, $[0, T]$. Each realization has a Fourier transform, $\tilde{\mathrm{x}}(t)$. Consider,

$$E\left[\frac{1}{T}|\tilde{\mathrm{x}}_T(t)|^2\right] \to \int_{-\infty}^{\infty} C_{\mathrm{x}}(\tau) e^{-i\omega\tau} d\tau = S_{\mathrm{x}}(\omega)$$

as $T \to \infty$. It can be shown that the average of Fourier transforms of realizations yields the power spectral density. (For details of this argument see Brown and Hwang 1997, ch. 2.7, p. 87-88).

Example 4.7 *Consider a random process with covariance function*

$$C_{\mathrm{x}}(\tau) = \sigma^2 e^{-|\tau|/T}.$$

What is the spectral density function?

By definition,

$$\begin{aligned}
S_{\mathrm{x}}(\omega) &= \int_{-\infty}^{\infty} C_{\mathrm{x}}(\tau) e^{-i\omega\tau} d\tau \\
&= \sigma^2 \int_{-\infty}^{\infty} e^{-|\tau|/T} e^{-i\omega\tau} d\tau \\
&= \sigma^2 \int_{-\infty}^{0} e^{(1/T - i\omega)\tau} d\tau + \sigma^2 \int_{0}^{\infty} e^{-(1/T + i\omega)\tau} d\tau \\
&= \sigma^2 \frac{1}{1/T - i\omega} e^{(1/T - i\omega)\tau}|_{-\infty}^{0} + \sigma^2 \frac{-1}{1/T + i\omega} e^{-(1/T + i\omega)\tau}|_{0}^{\infty} \\
&= \sigma^2 \frac{1}{1/T - i\omega} - \sigma^2 \frac{1}{1/T + i\omega}(-1) \\
&= \sigma^2 \frac{1/T + i\omega}{1/T^2 + \omega^2} + \sigma^2 \frac{1/T - i\omega}{1/T^2 + \omega^2} \\
&= \sigma^2 \frac{2/T}{1/T^2 + \omega^2} \qquad (4.13)
\end{aligned}$$

This function is illustrated in Maybeck (1979)'s Fig. 4.6. The middle panels show the autocorrelation function and its spectral density.

4.7 Special processes

4.7.1 Markov processes and sequences

A random process is called Markov if future statistics depend only on the present state and not on the past ones, i.e.

$$p(\boldsymbol{x}(t_i)|\boldsymbol{x}(\tau); \tau \leq t_{i-1}) = p(\boldsymbol{x}(t)|\boldsymbol{x}(t_i))$$

where $t_1 < t_2 < \ldots < t_i$. For discrete time processes (random sequences) we can say a random sequence is Markov if

$$p(\boldsymbol{x}_i|\boldsymbol{x}_k, k \leq i-1) = p(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}).$$

Markov processes are important because the solution to a first-order differential or difference equation forced by an independent process is Markov. For example, equations of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\boldsymbol{w}(t)$$

$$\mathbf{x}_k = \boldsymbol{\Phi}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\boldsymbol{w}_{k-1}.$$

As we shall see, these stochastic-dynamic equations define the true system to which the Kalman filter is applied. Additionally, note that for atmospheric dynamics, our forecast models could be described by

$$\dot{\mathbf{x}}(t) = F(\mathbf{x}, t)\mathbf{x}(t) + \mathbf{G}(\mathbf{x}, t)\boldsymbol{w}(t)$$

since the dynamics are nonlinear. Also, the function which maps model error into the model state, $\mathbf{G}$, may also be a function of the state.

4.7.2 Gaussian processes

$\mathbf{x}(t)$ in n-dimensions is Gaussian if for any $t_1, t_2, \ldots, t_N$, $p(\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_N))$ is jointly Gaussian. Let $\mathbf{z} = (\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_N))^{\mathrm{T}}$ then

$$p_{\mathbf{z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{(Nn/2)}|\mathbf{P}_{\mathbf{z}}|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathbf{z}})^{\mathrm{T}}\mathbf{P}_{\mathbf{z}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathbf{z}})].$$

Note that $\mathbf{z}$ is an Nn dimensional vector. $\mathbf{P}_{\mathbf{z}}$ is given by

$$\mathbf{P}_{\mathbf{z}} = E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^{\mathrm{T}}].$$

$\mathbf{P}_{\mathbf{z}}$ is an $Nn \times Nn$ matrix. $\mathbf{P}_{\mathbf{z}}$ can also be viewed as an $N \times N$ matrix with elements which are $n \times n$ matrices. The ij^{th} submatrix of $\mathbf{P}_{\mathbf{z}}$ is

$$(\mathbf{P}_{\mathbf{z}})_{ij} = E[(\mathbf{x}(t_i) - \boldsymbol{\mu}_{\mathbf{x}}(t_i))(\mathbf{x}(t_j) - \boldsymbol{\mu}_{\mathbf{x}}(t_j))^{\mathrm{T}}].$$

Gaussian processes have some important properties:

1. All statistics of a Gaussian random process are completely determined by its first and second order statistics.
2. Any linear function of a Gaussion random process is a Gaussian random process.
3. Uncorrelated Gaussian random processes are independent since $\mathbf{P}_{\mathbf{z}}$ becomes block diagonal.

4.7.3 Gauss-Markov processes

A stationary Gaussian process, $\mathrm{x}(t)$, with an exponential auto- correlation function is called Gauss-Markov. The auto-correlation function is then

$$\Gamma_{\mathrm{x}}(\tau) = \sigma^2 e^{-|\tau|/T}$$

with spectral density:

$$S_{\mathrm{x}}(\omega) = \frac{2/T}{1/T^2 + \omega^2}$$

where σ^2 is the mean square value and T is a time constant. Since

$$\lim_{\tau \to \infty} \Gamma_{\mathrm{x}}(\tau) = 0$$

the mean value, $E(\mathrm{x}(t))$ is 0. The Gauss-Markov process is important because it fits may physical processes (e.g. $x_{k+1} = fx_k + \epsilon_k, \epsilon_k$ is $\mathcal{N}(0, \sigma_{\mathrm{x}}^2)$) and because it has a simple mathematical description. As with all stationary Gaussian processes, specification of the autocorrelation function completely defines the process. Here is an example to illustrate this.

Example 4.8 *$\mathrm{x}(t)$ is Gauss-Markov with auto-correlation function,*

$$\Gamma_{\mathrm{x}}(\tau) = 100e^{-2|\tau|}.$$

What is the 3rd order p.d.f., $p_{\mathrm{x_1 x_2 x_3}}(x_1, x_2, x_3)$ where $x_1 = x(0), x_2 = x(1/2), x_3 = x(1)$?

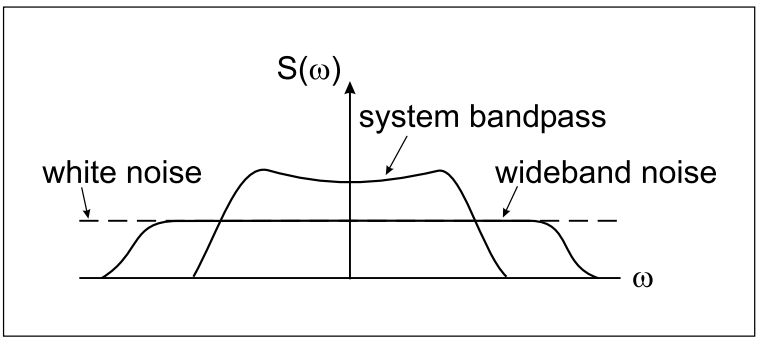

Figure 4.5: Wideband system noise. A system often has only a limited range of frequencies to which it can respond. If noise driving the system is wideband, then it is effectively white.

First note that $\lim_{\tau\to\infty}\Gamma_{\rm x}(\tau)=0$ *so that* $E(x(t))=0$. *Thus we can immediately write the third order p.d.f. as*

$$p_{\rm x_1x_2x_3}(x_1,x_2,x_3) \;=\; \frac{1}{(2\pi)^{(3/2)}|\mathbf{C}|^{1/2}}\exp[-\frac{1}{2}\mathbf{x}^{\rm T}\mathbf{C}^{-1}\mathbf{x}] \tag{4.14}$$

where $\mathbf{x}=(x_1,x_2,x_3)^{\rm T}$ *and*

$$\mathbf{C}=\begin{bmatrix} <x_1^2> & <x_2x_1> & <x_3x_1> \\ <x_1x_2> & <x_2^2> & <x_3x_2> \\ <x_1x_3> & <x_2x_3> & <x_3^2> \end{bmatrix}=\begin{bmatrix} 100 & 100e^{-1} & 100e^{-2} \\ 100e^{-1} & 100 & 100e^{-1} \\ 100e^{-2} & 100e^{-1} & 100 \end{bmatrix}. \tag{4.15}$$

4.7.4 White Noise

White noise is a stationary random process with a constant spectral density, $\Gamma(\tau)=C\delta(\tau)$. The spectral density is then $S(\omega)=C$, where

$$\int_{-\infty}^{\infty}\delta(\tau)d\tau=1$$

and

$$\delta(\tau)=\begin{cases}\infty & \tau=0\\ 0 & \tau\neq 0\end{cases}.$$

Such a process is physically unrealizable because it has infinite variance. Although not physical, white noise is useful mathematically for constructing (shaping) other processes by sending the white noise through a linear system. Also, the concept of band limited white noise is physically plausible. Band limited white noise has a constant spectral amplitude over a finite range of frequencies. For example, a system may have only a limited range of frequencies to which it can respond while the noise driving the system is constant over a wide band (extending over a broader range of frequencies). In this case, the noise is effectively white. This situation is illustrated in Fig. 4.5.

4.7.5 White Gaussian noise

$\mathbf{x}(t)$ is a white Gaussian process if for any $t_1,t_2,\ldots,t_N$, $(\mathbf{x}(t_1),\mathbf{x}(t_2),\ldots,\mathbf{x}(t_N))^{\rm T}$ are independent Gaussian random vectors. Thus the process is Gaussian in space, but white in time.

4.8 Spatially random fields

When a random variable is a fnction of time, it becomes a random process. When it is a function of space, it becomes a random field.

A random field is a function of the events ω in the probability space, Ω but also a function of spatial position, $\mathbf{r} \in \mathcal{R}^n$. For example, the random field $\mathbf{x}$ may be written as

$$\mathbf{x}(\mathbf{r}) = [\mathrm{x}_1(\mathbf{r}), \mathrm{x}_2(\mathbf{r}), \ldots, \mathrm{x}_n(\mathbf{r})]^{\mathrm{T}}.$$

A common example of a random vector field is the 3D velocity field, i.e.

$$\mathbf{x}(\mathbf{r}) = \begin{bmatrix} u(x,y,z) \\ v(x,y,z) \\ w(x,y,z) \end{bmatrix}.$$

We can define the probability distribution and densities for random fields using the same sort of extensions that were made for random processes. i.e. The probability distribution and density functions are

$$F_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) = P[\{\omega : \mathbf{x}(\mathbf{r}) \leq \boldsymbol{x}; \mathbf{r} \in \mathcal{R}^n\})$$
$$p_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) = \frac{\partial F_{\mathbf{x}}(\boldsymbol{x}, \boldsymbol{r})}{\partial \boldsymbol{x}}.$$

The mean value of a random field is given by

$$\boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}) = E[\boldsymbol{x}(\mathbf{r})] = \int_{-\infty}^{\infty} \boldsymbol{x} p_{\mathbf{x}}(\boldsymbol{x}, \mathbf{r}) d\boldsymbol{x} \tag{4.16}$$

and the auto-covariance matrix of a random spatial field is

$$\begin{aligned} \mathbf{C}_{\mathbf{x}}(\mathbf{r}_i, \mathbf{r}_j) = cov(\mathbf{x}(\mathbf{r}_i), \mathbf{x}(\mathbf{r}_j)) &= E\{[\mathbf{x}(\mathbf{r}_i) - \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_i)][\mathbf{x}(\mathbf{r}_j) - \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_j)]^{\mathrm{T}}\} \\ &= E\{\mathbf{x}(\mathbf{r}_i)\mathbf{x}(\mathbf{r}_j)^{\mathrm{T}}\} - \boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_i)\boldsymbol{\mu}_{\mathbf{x}}(\mathbf{r}_j)^{\mathrm{T}} \end{aligned} \tag{4.17}$$

for two points in space, $\mathbf{r}_i$ and $\mathbf{r}_j$.

For a scalar random field, $\mathrm{x}(\mathbf{r})$, the mean is a scalar field, $\mu(\mathbf{r})$ and the covariance matrix become a covariance function, $C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j)$. The variance is defined by the covariance function for $\mathbf{r}_i = \mathbf{r}_j$, i.e.

$$\sigma_{\mathrm{x}}^2(\mathbf{r}) = C_{\mathrm{x}}(\mathbf{r}, \mathbf{r}).$$

The spatial correlation function can then be defined as:

$$\rho_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j) = \frac{C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j)}{\sigma_{\mathrm{x}}(\mathbf{r}_i)\sigma_{\mathrm{x}}(\mathbf{r}_j)}.$$

A scalar random field is spatially uncorrelated when

$$C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} \sigma_{\mathrm{x}}^2(\mathbf{r}), & \text{for } \mathbf{r}_i = \mathbf{r}_j = \mathbf{r} \\ 0, & \text{otherwise} \end{cases}. \tag{4.18}$$

A spatially uncorrelated field is also called a **white** field.

Many of the concepts developed for random processes have analogies for random fields. For example, Markov processes correspond to Markov fields. Similarly, we can have Gaussian fields and white fields. The parallel concept to stationarity for random processes is that of *homogeneity* for random fields. For a homogeneous field, its mean value is independent of space and its covariance depends only on the vector distance between two points. Thus

$$\begin{aligned} \mu_{\mathrm{x}}(\mathbf{r}) &= \mu \\ C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j) &= C_{\mathrm{x}}(\mathbf{r}_i - \mathbf{r}_j) = C_{\mathrm{x}}(\mathbf{r}). \end{aligned} \tag{4.19}$$

A spatial field is *isotropic* if its covariance depends only on the scalar distance between two points, i.e.

$$C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j) = C_{\mathrm{x}}(|\mathbf{r}_i - \mathbf{r}_j|) = C_{\mathrm{x}}(r). \tag{4.20}$$

The spectral density function can be defined as the Fourier transform of the autocovariance function for homogeneous random fields. For real random fields,

$$C_{\mathrm{x}}(\mathbf{r}) = \int_{\mathcal{R}^n} \cos(\mathbf{w}^{\mathrm{T}}\mathbf{r}) \hat{C}_{\mathrm{x}}(\mathbf{w}) d\mathbf{w} \tag{4.21}$$

$$\hat{C}_{\mathrm{x}}(\mathbf{w}) = \frac{1}{(2\pi)^n} \int_{\mathcal{R}^n} \cos(\mathbf{w}^{\mathrm{T}}\mathbf{r}) C_{\mathrm{x}}(\mathbf{r}) d\mathbf{r}. \tag{4.22}$$

Using this, one can then determine criteria for which a function can be a covariance function. The necessary and sufficient condition for a continuous function $C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j)$ in $\mathcal{R}^n$ to be a covariance function is that it be positive semi-definite, that is,

$$\int_{\mathcal{R}^n} \int_{\mathcal{R}^n} C_{\mathrm{x}}(\mathbf{r}_i, \mathbf{r}_j) f(\mathbf{r}_i) f(\mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \geq 0 \tag{4.23}$$

for any function, $f(\mathbf{r})$. Using the spectral representation above, the condition for a continuous, symmetric function to be a covariance function is that its spectral function be positive semi-definite:

$$\hat{C}_{\mathrm{x}}(\mathbf{w}) \geq 0 \tag{4.24}$$

for $\mathbf{w} \in \mathcal{R}^n$. The special form for isotropic covariance functions in two-dimensions is discussed in Todling (1999).

The subject of random fields is a very big one and we have only summarized a few facts here. For more detail, see the books by Yaglom (1987) or Christakos (1992).

REFERENCES

1. Anderson, B. D. O., and J. B. Moore, 1979: *Optimal filtering.* Prentice-Hall, 357 pp.

2. Brown, R. G. and P. Y. C. Hwang, 1997: *Introduction to random signals and applied Kalman filtering.* John Wiley and sons, 484 pp.

3. Christakos, G., 1992: *Random field models in Earth Sciences.* Academic Press, 474 pp.

4. Gelb, A., 1974: *Applied optimal estimation.* The Analytic Sciences Corp., 374 pp.

5. Maybeck, P. S., 1979: *Stochastic models, estimation and control.* Vol. 1, Academic Press, 423 pp.

6. Papoulis, A., 1965: *Probability, random variables and stochastic processes.* McGraw-Hill. 583 pp.

7. Todling, R., 1999: *Estimation theory and foundations of atmospheric data assimilation.* DAO Office Note 1999-01.

8. Yaglom, A. M., 1987: *Correlation theory of stationary and related random functions, Vol. 1: Basic results.* Springer-Verlag, 526 pp.

4.9 Problem Set 4

1. (Todling ch. 2, #2) Calculate the power spectrum of the stationary processes having the following auto-covariance functions:

   (a) Gaussian pulse: $\Gamma(\tau) = \sigma^2 e^{-\tau^2/T^2}$

   (b) Damped cosine wave: $\Gamma(\tau) = \sigma^2 e^{-\beta|\tau|}\cos\omega_0\tau$

   (c) Triangular pulse:

$$\Gamma(\tau) = \begin{cases} 1-|\tau|, & \text{for} |\tau| \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

2. (Todling ch. 2, #3) The stationary process $x(t)$ has mean μ=const. and an autocorrelation function of the form

$$\Gamma(t, t+\tau) = \Gamma(\tau) = \sigma^2 e^{-\tau^2/T^2}$$

   Another process $y(T)$ is related to $x(t)$ by the deterministic equation

$$y(t) = ax(t) + b$$

   where a and b are known constants.
   (a) What is the auto-correlation function for $y(t)$?
   (b) What is the cross-correlation function $\Gamma_{xy}(\tau)$?

3. (Todling ch. 2, #4) Two random processes are defined by

$$\begin{aligned} x(t) &= asin(\omega t + \theta) \\ y(t) &= bsin(\omega t + \theta) \end{aligned}$$

   where θ is a random variable with uniform distribution between 0 and 2π, and ω is a known constant. The coefficients a and b are both normal random variables, N(0,σ^2) and are correlated with a correlation coefficient ρ. What is the cross-correlation function $\Gamma_{xy}(\tau)$? (Assume a and b are independent of θ.)

4. (Todling ch. 2, #5) Show that the following are admissible candidates for a covariance function:

   (a) $C_{\rm x}(\mathbf{r}) = a\delta(\mathbf{r})$, for $a > 0$ and $\mathbf{r} \in \mathcal{R}^n$,

(b) $C_{\mathrm{x}}(x, y) = C_{\mathrm{x}}(r = |x - y|) = \Pi \exp(-r^2)$, for $x, y \in \mathcal{R}$. (Hint: In this case, the proof can be obtained by either showing that (4.23) is true or showing that (4.24) is satisfied. Use (4.23) and expand $\exp(2xy)$ in Taylor series.)